Global variation in the HIV-1 V3 region

Brian Foley and Bette Korber

MS K710, Los Alamos National Laboratory, Los Alamos, NM 87545

Introduction

Due to the immunogenicity and functional importance of the V3 loop, there has been a great deal of interest in the V3 region of the envelope protein, resulting in a large international effort to obtain V3 region sequences. This section, which includes sequences taken from 967 individuals and complete references, provides an overview of the variation of sequences that span this region.

Sequences

To best summarize the spectrum of international HIV-1 variants, only one representative viral sequence was included per infected individual. A complete set of references accompanies the sequence alignments, and nomenclature was preserved from the original papers so individuals and isolates can be clearly identified. HIV1 was deleted from the sequence names in this section, as all sequences included here are HIV1. Included with the references when available are brief descriptions of critical features of the sequences. This includes the health status of the individual from whom the virus was derived, whether or not the virus was cultured, and the year the blood sample was taken.

All sequences are prefaced by a subtype association (see phylogenetic clustering below) and a two letter country code to identify the country that the individual resided in at the time that the blood sample was taken. If the person was a recent immigrant and this information was available, we included the country of origin in the references. The two letter code was developed for Internet (Copyright 1992, Lawrence H. Landwater and the Internet Society), and incorporated here based on a suggestion made by Dr. Francine McCutchan. The key to the country codes follows this introduction. Note that this key has been updated for 1995, with several country codes for eastern european nations added.

Sometimes only one viral sequence was available from a person: a clone from an isolate, or a direct sequence of PCR amplified peripheral blood DNA. For other individuals, up to 80 viral sequences from PCR amplified DNA or RNA from blood samples were available. Consequently, over 2000 sequences are represented by the 967 included in this section. When two sequences were available from a person, one of the two was randomly selected. When three or more sequences were generated from a person, all available sequences were aligned (without regard to different time points of sampling) and a consensus of the most common base found in each position in the alignment was generated. If there was a tie (e.g., 10 A's, 10 T's), the top base or amino acid in the alignment was used. If a set of sequences from two or more individuals was epidemiologically linked, and genetically very similar, only one sequence from the set was included, preferably the most recently infected. In the sequence description and references section, the short hand "PCR-direct, peripheral blood DNA" is used to signify that viral DNA was amplified from PBMCs, without culturing, and a single "direct" sequence was obtained from the amplification reaction products. The short hand "Consensus, PCR-clones, peripheral blood DNA" signifies that viral DNA was amplified from PBMCs and a set of clones was generated and sequenced from the PCR amplification products. The cloned sequences were aligned and a consensus was generated. In a handful of cases, a particular gp160 clone from an isolate was shown to be expressed and functional using a vaccinia virus T7 expression system. In these cases, the clone rather than the consensus of all sequences from a particular individual is included.

Phylogenetic clustering

Sequences have been organized according to the phylogenetic subtype association (A-H) of their envelope V3 regions only. The original sequence subtype (A-H) designations were defined based on the phylogenetic relationships determined by using both gag and env genes (when possible), are

approximately genetically equidistant in envelope, and have multiple members. The phylogenetic subtype designations and associations have generally been adopted by the HIV research community, and are now often presented with the publication of new sequences. We have either determined the subtype designations here, if not specified in the original manuscript, or else confirmed the subtype designations of the original manuscripts, and then used the subtypes to organize this section. Generally, confirmations were done by aligning a set HIV-1 V3 region sequences with longer env gene sequences (Part IIIC) that have clear subtype associations, and then using parsimony or neighbor joining trees to determine associations. Some of the shorter gene fragments from this region were given a subtype designation based on Hamming distances, using the similarity function of the MASE program (Faulkner DV, and Jurka J. TIBS 13:321-322 (1988)); these sequences have ".sh" appended to their name to indicate that they were too short for phylogenetic analysis. Parsimony trees were generated using PAUP (David Swofford, Illinois Natural History Survey), and neighbor-joining trees were generated with Kimura distances using PHYLIP (Joseph Felsenstein, University of Washington). All available nucleotide sequence information was used for phylogenetic analysis; longer protein sequences were trimmed to be approximately the same length as the majority of the PCR fragments in this region, for the purposes of presentation. Some sequences were difficult to classify, and are included in the "U", or unclassified, section. In addition, it has recently been noted that recombination between HIV-1 occurs when an individual is infected with more than one strain. A meeting was held in Santa Fe, New Mexico in October, 1995 to discuss the implications of recombination and methods for detecting recombinant sequences. Because inter-subtype as well as intra-subtype recombination is known to occur, the subtype designations reported in this section should be interpretted only as pertaining to the V3 region of the envelope gene. For example HIV-1 MAL from Zaire, is known to be recombinant between subtypes A and D, with the V3 loop of env resembling subtype D. D_ZR-MAL is still listed with other subtype D sequences in this study, but may be moved to the U (uncertain) group in the future.

The set of sequences used to help resolve subtype associations included at least two sequences from each subtype (A-H), plus a simian immunodeficiency virus outgroup sequence. The sequences were selected based on being "typical" of the subtype they represent based on phylogenetic analysis. The set has changed as more sequences have accumulated. Thus not all subtype designations were based on the same reference set.

Limitations of phylogenetic analyses

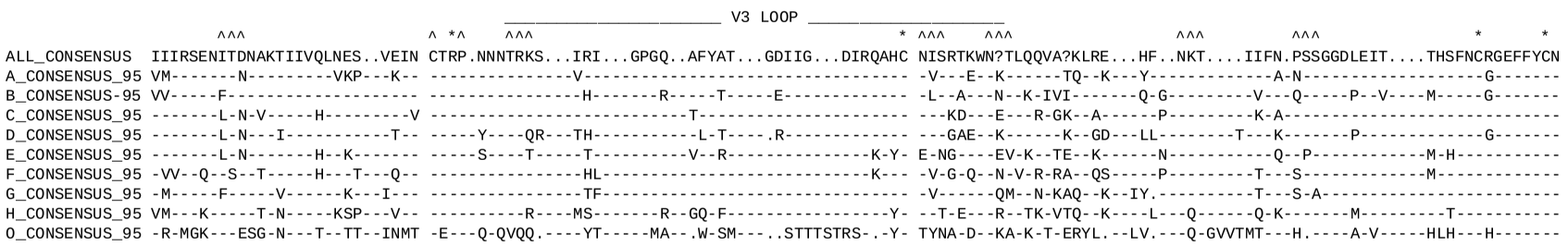
Most of the PCR derived sequences contain a sub-optimal length for phylogenetic analyses, given the level of variability in this region – typically on the order of 250 to 300 nucleotides. Due to this limitation, some of the classifications in this section are uncertain and are our best estimate given the available information. Control studies were performed to compare the phylogenetic clustering of the V3 region using available longer sequences, however, and these studies indicate that our subtype designations based on the V3 region are generally reliable. For 146 sequences, we had an approximately 700 base region of env available representing all of the subtypes A-H. (The limitation in length was due to including the H subtype sequences, which did not cover all of gp120.) After removing positions in the alignment which included gaps, 519 bases were left. When a 298 base V3 region fragment was excised from this set, and neighbor joining trees were constructed using both the 519 base and 298 base long sequences, the phylogenetic subtype designations were consistent in each case. Further, when a subset of longer gp120 sequences was analyzed (92 of the 146), including 935 bases after removing positions in the alignment which included gaps, the subtype designations were again clear in neighbor-joining trees. This indicates that the limited V3 region PCR fragments, which include more than the V3-loop, are generally able to serve as a reliable basis for subtype determination.

Without detailed analysis, genetic recombination between subtypes may obscure phylogenetic relationships between sequences. A characteristic of recombination is an indeterminate place in phylogenetic analyses, and some of the "Uncertain" category sequences may prove to be recombinant genomes upon further inspection. Also, while a subtype designation based on a gene or gene fragment may be correct, recombination events outside the region examined may have occurred. Therefore, care should be taken to not overinterpret the subtype designations. If one is to discuss the subtype

designations of viral isolates based on the data presented here, they should be refer to the designation as “B-like over V3 loop region,” rather than as “subtype B”.

Limitations of V3 amino acid consensus sequences

The V3 amino acid consensus sequences generated for each subtype have interesting features; however, one should be wary about assuming that any of the consensus sequences may broadly represent their subtype. Certainly many V3 loop variants in each of the subtypes are extremely divergent from the consensus sequences. These divergent forms may have very different biological and immunological characteristics from viruses which are similar to the consensus. Additionally, because of the relatively small sample size of most of the subtypes, consensus sequences can be dominated by a small group of highly similar sequences, which may in turn be a sampling artifact. Hence, these consensus sequences are “evolving” as new sequences from each subtype become available.

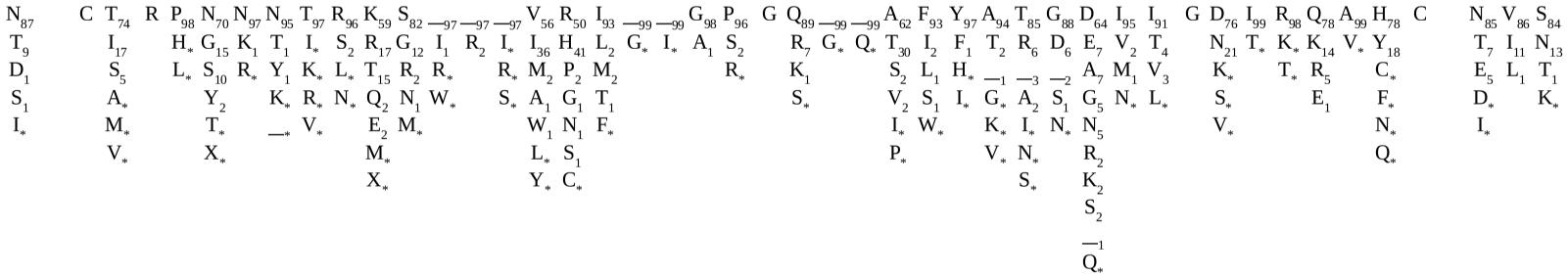


Subtype consensus sequences. This V3 region alignment shows a consensus sequence generated for each of the nine subtypes. The nine subtype consensus sequences indicate the most common amino acid found in each position among the sequences associated with each subtype. The sequences are aligned to a consensus based on the most common amino acid in the subtype consensus sequences, which approximates a "global" consensus. It was generated in this way (rather than by using all 952 sequences) to avoid over-representation of the B subtype, which has by far the largest number of available sequences. As is the convention in this compendium, a dash (-) indicates concurrence with the top sequence in the alignment; a period (.) indicates a deletion. The carets show where the N-linked glycosylation sites are found in the consensus. The V3 loop is set off from the surrounding sequence by a space on either side to facilitate viewing. Interesting features of the consensus alignment are: 1) Only in the B subtype is GPGR the most common tip of the V3 loop; globally, GPGQ is more prevalent. 2) A highly conserved N-linked glycosylation site is constitutively absent in the C subtype, proximal to the first cysteine (C) in loop. 3) The D subtype consensus has 34 amino acids from cysteine (C) to cysteine (C) rather than the more common 35; at the point where the deletion occurs, it is not uncommon to find insertions of 2 to 4 amino acids, as can be observed in the sequence alignments. 4) The D subtype has two arginine (R) residues in the V3 loop that are uncharacteristic relative to the other consensus sequences; positively charged amino acids in these positions may result in a syncytia inducing, non-monocytropic phenotype (Fouchier RAM, et al., *J. Virol.* **66**:3183–3187 (1992)). 5) A higher degree of variation is seen in the region just downstream of the V3 loop than within it. This difference is also observed internally among the sequences of the different subtypes. 6) The A, C, G and H consensus sequences have very similar V3 loop sequences.

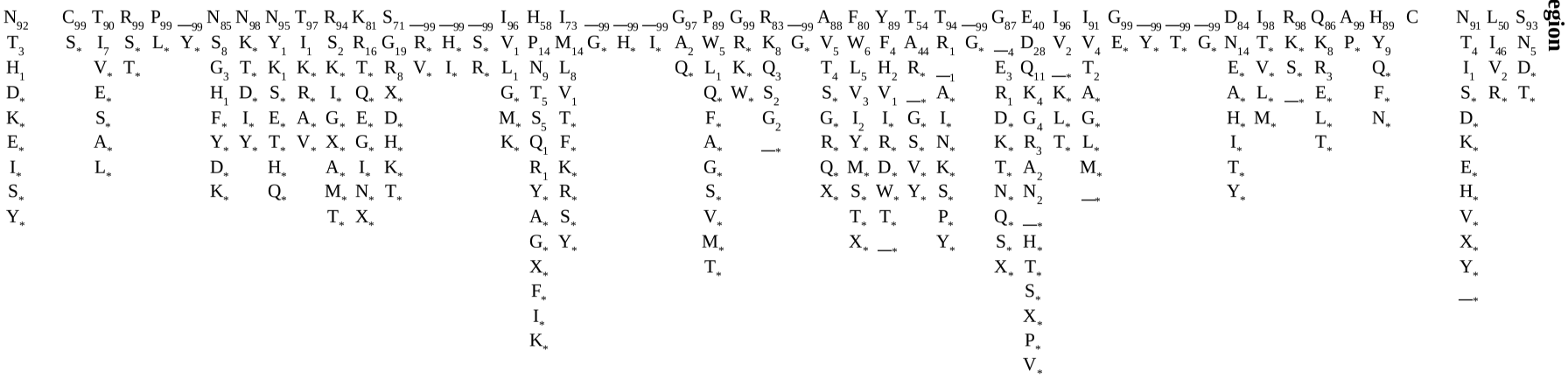
V3 Loop Amino Acids

The following pages present amino acid alignments of the V3 loop, arranged by phylogenetic subtype. For each subtype, the number of sequences used to construct the alignment is indicated. The top line in each alignment represents the consensus sequence for that subtype, where consensus simply means the most common amino acid found in each position among the sequences of the given subtype. The subscripts record the frequency with which that amino acid is observed at that location among members of the subtype. An amino acid which is conserved 100% is shown with no subscript. Directly beneath the most common amino acid in each position are the other amino acids observed in that position, listed from most common to least common. An asterisk (*) subscript means less than 1% of the sequences had the indicated amino acid at that location. A dash (-) indicates a gap inserted to maintain the alignment.

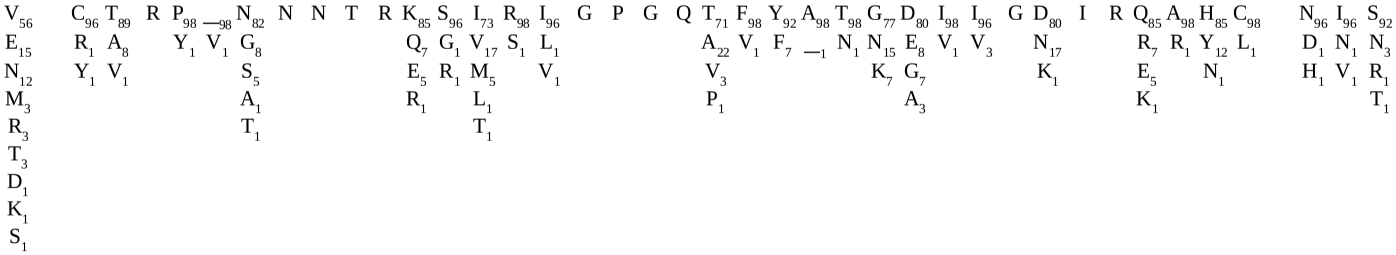
A subtype (139 sequences)



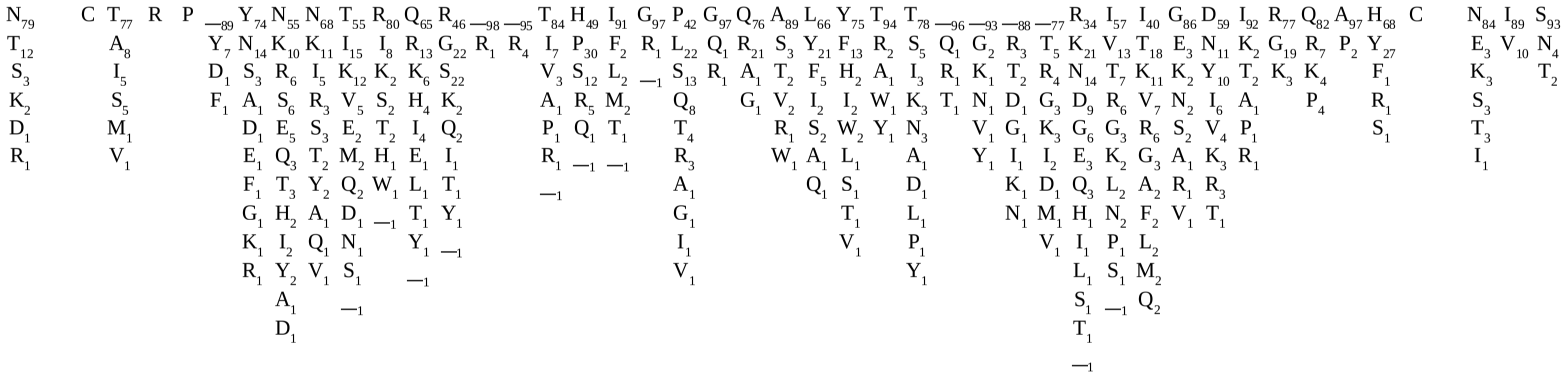
B subtype (519 sequences)



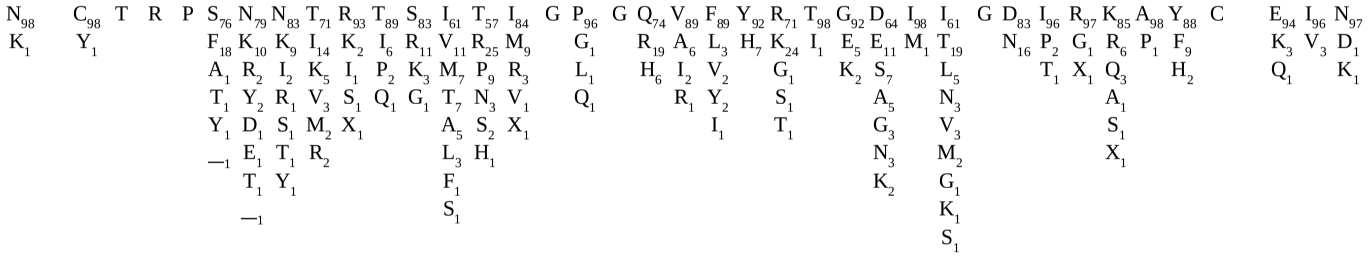
C subtype (57 sequences)



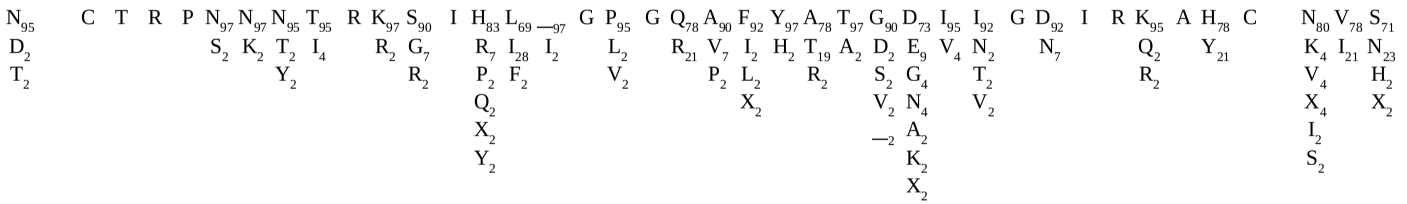
D subtype (89 sequences)



E subtype (77 sequences)



F subtype (42 sequences)



G subtype (11 sequences)

N$_{45}$ C T$_{72}$ R P N N N T R K S I T$_{27}$ F$_{45}$ G$_{90}$ P$_{90}$ G Q$_{90}$ A$_{90}$ F$_{72}$ Y A$_{90}$ T G$_{90}$ D$_{45}$ I I G D$_{72}$ I R$_{90}$ Q A H$_{81}$ C N V$_{90}$ S$_{81}$
I$_{27}$ I$_{18}$ H$_{18}$ I$_{27}$ A$_{9}$ T$_{9}$ R$_{9}$ V$_{9}$ L$_{18}$ T$_{9}$ S$_{9}$ A$_{27}$ N$_{27}$ K$_{9}$ Y$_{18}$ I$_{9}$ N$_{18}$
T$_{18}$ V$_{9}$ N$_{18}$ L$_{27}$ I$_{9}$ E$_{18}$
V$_{9}$ K$_{9}$ N$_{9}$
P$_{9}$
R$_{9}$
S$_{9}$

H subtype (2 sequences)

N C T R P N N N T R K$_{50}$ S I$_{50}$ R$_{50}$ I G I$_{50}$ G Q$_{50}$ A$_{50}$ F$_{50}$ T$_{50}$ A$_{50}$ H$_{50}$ A$_{50}$ I$_{50}$ G A$_{50}$ I I G D I R K$_{50}$ A H$_{50}$ C N I S$_{50}$
R$_{50}$ M$_{50}$ S$_{50}$ P$_{50}$ R$_{50}$ G$_{50}$ Q$_{50}$ Y$_{50}$ F$_{50}$ T$_{50}$ —$_{50}$ —$_{50}$ D$_{50}$ Q$_{50}$ Y$_{50}$ T$_{50}$

O subtype (3 sequences)

T$_{66}$ C E$_{66}$ R P$_{66}$ —$_{66}$ G$_{33}$ A$_{33}$ D$_{33}$ I$_{33}$ Q$_{66}$ D$_{33}$ —$_{66}$ I$_{66}$ M$_{33}$ A$_{33}$ G P M A$_{66}$ W Y$_{66}$ S M —$_{66}$ L$_{33}$ L$_{33}$ G$_{33}$ I$_{33}$ G$_{33}$ G$_{33}$ K$_{33}$ A$_{33}$ D$_{33}$ N$_{33}$ R$_{66}$ S$_{66}$ R$_{66}$ A$_{33}$ A Y C K$_{33}$ Y N$_{66}$
A$_{33}$ I$_{33}$ E$_{33}$ G$_{33}$ I$_{33}$ I$_{33}$ E$_{33}$ T$_{33}$ I$_{33}$ E$_{33}$ K$_{33}$ M$_{33}$ R$_{33}$ I$_{33}$ R$_{33}$ R$_{33}$ A$_{33}$ T$_{33}$ S$_{33}$ K$_{33}$ R$_{33}$ N$_{33}$ N$_{33}$ G$_{33}$ G$_{33}$ P$_{33}$ S$_{33}$ A$_{33}$ —$_{33}$ V$_{33}$ N$_{33}$ S$_{33}$
Q$_{33}$ N$_{33}$ Q$_{33}$ V$_{33}$ Q$_{33}$ Y$_{33}$ T$_{33}$ —$_{33}$ —$_{33}$ —$_{33}$ —$_{33}$ S$_{33}$ T$_{33}$ T$_{33}$ T$_{33}$ S$_{33}$ T$_{33}$ —$_{33}$ T$_{33}$

U subtype (13 sequences)

N$_{76}$ C T R P N$_{76}$ N$_{92}$ N$_{69}$ T$_{76}$ R$_{84}$ K$_{53}$ S$_{76}$ —$_{84}$ —$_{84}$ I$_{69}$ R$_{69}$ I$_{69}$ G P$_{92}$ G Q$_{61}$ A$_{61}$ F$_{84}$ Y$_{84}$ A$_{69}$ T$_{76}$ G$_{61}$ D$_{69}$ I$_{76}$ I$_{69}$ G D$_{76}$ I$_{92}$ R$_{92}$ Q$_{76}$ A H$_{61}$ C N$_{69}$ I$_{76}$ S$_{46}$
S$_{7}$ G$_{15}$ S$_{7}$ D$_{7}$ I$_{7}$ K$_{7}$ R$_{23}$ R$_{15}$ Q$_{15}$ R$_{7}$ V$_{15}$ H$_{23}$ F$_{15}$ S$_{7}$ R$_{30}$ T$_{23}$ I$_{15}$ F$_{15}$ T$_{23}$ I$_{7}$ S$_{15}$ A$_{7}$ M$_{7}$ T$_{15}$ N$_{15}$ —$_{7}$ —$_{7}$ K$_{23}$ Y$_{30}$ D$_{7}$ V$_{23}$ N$_{38}$
T$_{7}$ S$_{7}$ I$_{7}$ K$_{7}$ N$_{7}$ I$_{15}$ G$_{7}$ S$_{7}$ M$_{7}$ N$_{7}$ L$_{7}$ K$_{7}$ V$_{15}$ —$_{7}$ K$_{7}$ —$_{15}$ G$_{7}$ V$_{7}$ K$_{7}$ —$_{7}$ Q$_{7}$ I$_{7}$ T$_{15}$
V$_{7}$ T$_{7}$ R$_{7}$ T$_{7}$ T$_{7}$ V$_{7}$ —$_{7}$ N$_{7}$ K$_{7}$ —$_{7}$ —$_{7}$ S$_{7}$
Y$_{7}$ —$_{7}$ T$_{7}$

Summary of variations in the tetrameric tip of the V3 loop. This table is a tally of the different tetramers observed in the 976 individuals analyzed. This tip is thought to form a turn, and is the focal point of the potent neutralizing antibody epitopes that have been mapped to the V3 loop, as well as of T cell epitopes. Each column shows the number of occurrences of a given tetramer in either the entire 976 sequences (combined), or in subsets consisting of subtypes A–O, and the unclassified sequences (U). Underneath the column heading is the number of sequences in each category. The most common form found in each subtype is highlighted in bold lettering. In the B subtype, GPGR is the predominant form, however globally GPGQ is more common.

	Combined	A	B	C	D	E	F	G	H	O	U
	976	138	519	57	88	94	42	11	2	12	13
GPGR	424	9	369		16	19	7				4
GPGQ	324	119	15	57	21	62	33	9	1		7
GPGK	45	2	42								1
GWGR	28		28								
GLGQ	19				19						
GPGS	13	1	12								
GSGQ	16	3			12					1	
APGR	14	1	13								
GPGG	11		11								
GQGQ	8		1		7						
GPGH	8					8					
GLGR	8		6			1	1				
GPMA	6									6	
GQGR	5		2		1	2					
GTGQ	4				4						
GRGQ	4	1			3						
GVGR	4		2		1		1				
GSGR	3	1	2								
GGGQ	3					2					
GFGR	3		3								
GPLS	2									2	
GGGR	2		2								
APGQ	2	1						1			
GPMS	1									1	
GLGS	1		1								
GTGR	1							1			
GTGG	1		1								
GPMR	1									1	
GPKR	1		1								
GPGA	1				1						
GLRQ	1				1						
AQGR	1				1						
GQRK	1		1								
GPLR	1									1	
GPLA	1									1	
GARR	1		1								
GAGR	1		1								
APGS	1		1								
GMGR	1		1								
GPRR	1		1								
GPWG	1		1								
GIGQ	1				1						
GIGR	1								1		

Summary of variations in the octameric tip of the V3 loop. This table is a tally of the different octamers observed in the 967 individuals analyzed. This table is structured the same as the tetramer table on the previous pages. Amino acid changes proximal to the tip can influence the specificity of anti-V3 neutralizing antibodies. The 189 forms that were found only once in the data set are not shown here, to save space, and are summarized in a row labeled "unique."

	Combined	A	B	C	D	E	F	G	H	O	U
	967	138	519	57	88	94	42	11	2	3	13
HIGPGRAF	162		158		2		2				
RIGPGQTF	69	27		40						2	
RIGPGQAF	47	33		12							2
PIGPGRAF	43		43								
HIGPGQAF	37	31	2				4				
NIGPGRAF	33		33								
TIGPGQVF	26					26					
HLGPGQAF	26						26				
SIGPGRAF	20		20								
RIGPGQVF	16			2		14					
HIGPGKAF	14		14								
PIGLGQAL	13				13						
HIAPGKAF	12		12								
HMGWGRAF	11		11								
TIGPGRAF	11		11								
RIGPGRVF	10					10					
HIGPGRAY	9				9						
HIGPGQAL	8				8						
NMGPGRAF	7		7								
HIGPGSAF	6		6								
PLGPGQAW	6		6								
PMGPGRAF	6		6								
HIGPGQTF	5	5									
HIGPGRAV	5		5								
HIGSGQAL	5				5						
HIGPGRTF	5		5								
HLGPGQAW	5		5								
HLGWGRAF	5		5								
HMGPGKAF	5		5								
HMGWGRTF	5		5								
HMGPGRAF	4		4								
TMGPGQVF	4					4					
PMGPGKAF	4		4								
TRGPGHVF	4					4					
PIGPGKAF	3		3								
QIGPGRAF	3		3								
HIGPGRAI	3		3								
HIGPGGAF	3		3								
NIGPGRAW	3		3								
NIGPGQVF	3					3					
PIGPGQAF	3					3					
PIGPGQVF	3					3					
TMGPGHVF	3					3					
HIGPGRAF	3	3									
GIGPGQTF	3	3									

	Combined	A	B	C	D	E	F	G	H	O	U
	967	138	519	57	88	94	42	11	2	3	13
TMGPGRVW	3		3								
YIGPGRAV	3		3								
HIGSGQAY	3				3						
HLGPGQAF	2	2									
PIGPGRAW	2		2								
NIGPGQAF	2	2									
PIGPGQAF	2	2									
RIGPGQSF	2	2									
RIGPGQVF	2	2									
AIGPGRTV	2		2								
HIGGGRTL	2		2								
HIGPGRAW	2		2								
HIGPGRVF	2		2								
RFGPGQAF	2										2
HLGFGRAL	2		2								
HLGPGGAF	2		2								
HLGPGRAW	2		2								
HMGPGGAF	2		2								
HMGPGGAL	2		2								
HMGPGKTF	2		2								
HIGPGQAI	2				2						
HIGPGQAY	2				2						
HIGTGQAL	2				2						
HMGPGRAL	2				2						
PIGLGQAY	2				2						
PIGRGQAL	2				2						
PIGSGQAL	2				2						
RIGPGQAL	2				2						
SIGLGQAL	2				2						
SIGQGQAL	2				2						
HIGGGQAY	2					2					
PIGPGQVL	2					2					
SIGQGRVL	2					2					
TIGPGQIF	2					2					
TIGPGRRF	2					2					
TIGPGRVF	2					2					
TIGPGRVY	2					2					
QRGPGRAF	2		2								
SIGPGRAW	2		2								
SIGPGRVW	2		2								
TMGPGKVF	2		2								
TMGPGRVL	2		2								
TMGPGRVY	2		2								
YIGPGRAF	2		2								
TMGPGRVF	2					2					
UNIQUE	189	26	93	3	26	8	10	11	2	3	7

COUNTRY CODES

It is becoming increasingly useful to name viral isolates and samples with a country code. The following code was captured from Internet files:

gopher://kupe.itu.ch/11/.1/itudoc/public/gophermenus/.1/.un/.edicore/.wp4/.sept95/.rdocs95/.28024
gopher://kupe.itu.ch/00/.1/ITU-Databases/.1/CtryCodes/.full-list.txt

for ISO two-letter codes and for ITU three-letter codes, respectively.

This is a list based on the International Organization for Standardization (ISO) 3166:1993 standard, updated from a list prepared by Mark Horton. Note that the original standard has this same information sorted into about 6 different orders, both in English and French, therefore this is an abbreviated version not to be taken as the entire standard. While it has been checked against the standard, it may possibly contain errors; the standard and registration newsletters should be verified for any critical application.

This copy has been updated and is believed to be current through September 1995.

Table of Country Codes from ISO 3166

Country	A 2	A 3	Number
AFGHANISTAN	AF	AFG	004
ALBANIA	AL	ALB	008
ALGERIA	DZ	DZA	012
AMERICAN SAMOA	AS	ASM	016
ANDORRA	AD	AND	020
ANGOLA	AO	AGO	024
ANGUILLA	AI	AIA	660
ANTARCTICA	AQ	ATA	010
ANTIGUA AND BARBUDA	AG	ATG	028
ARGENTINA	AR	ARG	032
ARUBA	AW	ABW	533
AUSTRALIA	AU	AUS	036
AUSTRIA	AT	AUT	040
AZERBAIJAN	AZ	AZE	031
BAHAMAS	BS	BHS	044
BAHRAIN	BH	BHR	048
BANGLADESH	BD	BGD	050
BARBADOS	BB	BRB	052
BELGIUM	BE	BEL	056
BELIZE	BZ	BLZ	084
BENIN	BJ	BEN	204
BERMUDA	BM	BMU	060
BHUTAN	BT	BTN	064
BOLIVIA	BO	BOL	068
BOTSWANA	BW	BWA	072
BOUVET ISLAND	BV	BVT	074
BOSNIA AND HERZEGOVINA	BA	BIH	070
BRAZIL	BR	BRA	076
BRITISH INDIAN OCEAN TERRITORY	IO	IOT	086
BRUNEI DARUSSALAM	BN	BRN	096
BULGARIA	BG	BGR	100
BURKINA FASO	BF	BFA	854
BURUNDI	BI	BDI	108

BYELORUSSIAN SSR	BY	BYS	112
CAMBODIA	KH	KHM	116
CAMEROON	CM	CMR	120
CANADA	CA	CAN	124
CAPE VERDE	CV	CPV	132
CAYMAN ISLANDS	KY	CYM	136
CENTRAL AFRICAN REPUBLIC	CF	CAF	140
CHAD	TD	TCD	148
CHILE	CL	CHL	152
CHINA	CN	CHN	156
CHRISTMAS ISLAND	CX	CXR	162
COCOS (KEELING) ISLANDS	CC	CCK	166
COLOMBIA	CO	COL	170
COMOROS	KM	COM	174
CONGO	CG	COG	178
COOK ISLANDS	CK	COK	184
COSTA RICA	CR	CRI	188
COTE D'IVOIRE	CI	CIV	384
CROATIA	HR	HRV	191
CUBA	CU	CUB	192
CYPRUS	CY	CYP	196
CZECH REPUBLIC	CZ	CZE	203
DENMARK	DK	DNK	208
DJIBOUTI	DJ	DJI	262
DOMINICA	DM	DMA	212
DOMINICAN REPUBLIC	DO	DOM	214
EAST TIMOR	TP	TMP	626
ECUADOR	EC	ECU	218
EGYPT	EG	EGY	818
EL SALVADOR	SV	SLV	222
EQUATORIAL GUINEA	GQ	GNQ	226
ERITREA	ER	ERI	232
ESTONIA	EE	EST	233
ETHIOPIA	ET	ETH	230
FALKLAND ISLANDS (MALVINAS)	FK	FLK	238
FAROE ISLANDS	FO	FRO	234
FIJI	FJ	FJI	242
FINLAND	FI	FIN	246
FRANCE	FR	FRA	250
FRANCE, METROPOLITAN	FX		249
FRENCH GUIANA	GF	GUF	254
FRENCH POLYNESIA	PF	PYF	258
FRENCH SOUTHERN TERRITORIES	TF	ATF	260
GABON	GA	GAB	266
GAMBIA	GM	GMB	270
GEORGIA	GE	GEO	268
GERMANY	DE	DEU	276
GHANA	GH	GHA	288
GIBRALTAR	GI	GIB	292
GREECE	GR	GRC	300
GREENLAND	GL	GRL	304
GRENADA	GD	GRD	308
GUADELOUPE	GP	GLP	312
GUAM	GU	GUM	316

GUATEMALA	GT	GTM	320
GUINEA	GN	GIN	324
GUINEA-BISSAU	GW	GNB	624
GUYANA	GY	GUY	328
HAITI	HT	HTI	332
HEARD AND MCDONALD ISLANDS	HM	HMD	334
HONDURAS	HN	HND	340
HONG KONG	HK	HKG	344
HUNGARY	HU	HUN	348
ICELAND	IS	ISL	352
INDIA	IN	IND	356
INDONESIA	ID	IDN	360
IRAN (ISLAMIC REPUBLIC OF)	IR	IRN	364
IRAQ	IQ	IRQ	368
IRELAND	IE	IRL	372
ISRAEL	IL	ISR	376
ITALY	IT	ITA	380
JAMAICA	JM	JAM	388
JAPAN	JP	JPN	392
JORDAN	JO	JOR	400
KAZAKHSTAN	KZ	KAZ	398
KENYA	KE	KEN	404
KIRIBATI	KI	KIR	296
KOREA, DEMOCRATIC PEOPLE'S REPUBLIC OF	KP	PRK	408
KOREA, REPUBLIC OF	KR	KOR	410
KUWAIT	KW	KWT	414
KYRGYZSTAN	KG	KGZ	417
LAO PEOPLE'S DEMOCRATIC REPUBLIC	LA	LAO	418
LATVIA	LV	LVA	428
LEBANON	LB	LBN	422
LESOTHO	LS	LSO	426
LIBERIA	LR	LBR	430
LIBYAN ARAB JAMAHIRIYA	LY	LBY	434
LIECHTENSTEIN	LI	LIE	438
LITHUANIA	LT	LTU	440
LUXEMBOURG	LU	LUX	442
MACAU	MO	MAC	446
MADAGASCAR	MG	MDG	450
MALAWI	MW	MWI	454
MALAYSIA	MY	MYS	458
MALDIVES	MV	MDV	462
MALI	ML	MLI	466
MALTA	MT	MLT	470
MARSHALL ISLANDS	MH	MHL	584
MARTINIQUE	MQ	MTQ	474
MAURITANIA	MR	MRT	478
MAURITIUS	MU	MUS	480
MAYOTTE	YT	MYT	175
MEXICO	MX	MEX	484
MICRONESIA	FM	FSM	583
MOLDOVA	MD	MDA	498
MONACO	MC	MCO	492
MONGOLIA	MN	MNG	496
MONTSERRAT	MS	MSR	500

MOROCCO	MA	MAR	504
MOZAMBIQUE	MZ	MOZ	508
MYANMAR	MM	MMR	104
NAMIBIA	NA	NAM	516
NAURU	NR	NRU	520
NEPAL	NP	NPL	524
NETHERLANDS	NL	NLD	528
NETHERLANDS ANTILLES	AN	ANT	532
NEUTRAL ZONE	NT	NTZ	536
NEW CALEDONIA	NC	NCL	540
NEW ZEALAND	NZ	NZL	554
NICARAGUA	NI	NIC	558
NIGER	NE	NER	562
NIGERIA	NG	NGA	566
NIUE	NU	NIU	570
NORFOLK ISLAND	NF	NFK	574
NORTHERN MARIANA ISLANDS	MP	MNP	580
NORWAY	NO	NOR	578
OMAN	OM	OMN	512
PAKISTAN	PK	PAK	586
PALAU	PW	PLW	585
PANAMA	PA	PAN	590
PAPUA NEW GUINEA	PG	PNG	598
PARAGUAY	PY	PRY	600
PERU	PE	PER	604
PHILIPPINES	PH	PHL	608
PITCAIRN	PN	PCN	612
POLAND	PL	POL	616
PORTUGAL	PT	PRT	620
PUERTO RICO	PR	PRI	630
QATAR	QA	QAT	634
REUNION	RE	REU	638
ROMANIA	RO	ROM	642
RUSSIAN FEDERATION	U	RUS	643
RWANDA	RW	RWA	646
ST. HELENA	SH	SHN	654
SAINT KITTS AND NEVIS	KN	KNA	659
SAINT LUCIA	LC	LCA	662
ST. PIERRE AND MIQUELON	PM	SPM	666
SAINT VINCENT AND THE GRENADINES	VC	VCT	670
SAMOA	WS	WSM	882
SAN MARINO	SM	SMR	674
SAO TOME AND PRINCIPE	ST	STP	678
SAUDI ARABIA	SA	SAU	682
SENEGAL	SN	SEN	686
SEYCHELLES	SC	SYC	690
SIERRA LEONE	SL	SLE	694
SINGAPORE	SG	SGP	702
SLOVAKIA	SK	SVK	703
SLOVENIA	SI	SVN	705
SOLOMON ISLANDS	SB	SLB	090
SOMALIA	SO	SOM	706
SOUTH AFRICA	ZA	ZAF	710
SOUTH GEORGIA AND THE SOUTH SANDWICH ILANDS	GS		239

SPAIN	ES	ESP	724
SRI LANKA	LK	LKA	144
SUDAN	SD	SDN	736
SURINAME	SR	SUR	740
SVALBARD AND JAN MAYEN ISLANDS	SJ	SJM	744
SWAZILAND	SZ	SWZ	748
SWEDEN	SE	SWE	752
SWITZERLAND	CH	CHE	756
SYRIAN ARAB REPUBLIC	SY	SYR	760
TAIWAN, PROVINCE OF CHINA	TW	TWN	158
TAJIKISTAN	TJ	TJK	762
TANZANIA, UNITED REPUBLIC OF	TZ	TZA	834
THAILAND	TH	THA	764
TOGO	TG	TGO	768
TOKELAU	TK	TKL	772
TONGA	TO	TON	776
TRINIDAD AND TOBAGO	TT	TTO	780
TUNISIA	TN	TUN	788
TURKEY	TR	TUR	792
TURKMENISTAN	TM	TKM	795
TURKS AND CAICOS ISLANDS	TC	TCA	796
TUVALU	TV	TUV	798
UGANDA	UG	UGA	800
UKRAINE	UA	UKR	804
UNITED ARAB EMIRATES	AE	ARE	784
UNITED KINGDOM	GB	GBR	826
UNITED STATES	US	USA	840
UNITED STATES MINOR OUTLYING ISLANDS	UM	UMI	581
URUGUAY	UY	URY	858
UZBEKISTAN	UZ	UZB	860
VANUATU	VU	VUT	548
VATICAN CITY STATE (HOLY SEE)	VA	VAT	336
VENEZUELA	VE	VEN	862
VIET NAM	VN	VNM	704
VIRGIN ISLANDS (BRITISH)	VG	VGB	092
VIRGIN ISLANDS (U.S.)	VI	VIR	850
WALLIS AND FUTUNA ISLANDS	WF	WLF	876
WESTERN SAHARA	EH	ESH	732
YEMEN, REPUBLIC O	YE	YEM	887
YUGOSLAVIA	YU	YUG	890
ZAIRE	ZR	ZAR	180
ZAMBIA	ZM	ZMB	894
ZIMBABWE	ZW	ZWE	716

V3 LOOP

```
                                                                 ______________________ V3 LOOP ______________________
                         ^^^                    ^ *^    ^^^                                                  * ^^^    ^^^                 ^^^         ^^^                  *      *
A_CONSENSUS_95   VMIRSENITNNAKTIIVQLVKP..VKIN CTRP.NNNTRKS...VRI...GPGQ..AFYAT...GDIIG...DIRQAHC NVSRTEWNKTLQQVATQLRK...YF..NKT....IIFA.NSSGGDLEIT....THSFNCGGEFFYCN
A_KE.KEN985      -V---------V--------E-..-I-- ----.-------...I--...----..T---...-E---...N------ ----SK--N---K------E...H-
A_NG.NI          -V-----------------AN-..---- ----.------G...-H-...----..-----...-----...------- ---K-------H--V---KT...--K.-T-...----.-PL---V---....--------------
A_RW.564C        -R--A----D-----M-H-NT-..-Q-- ----.-------...---...----..T----...-E---...N----Y- -LN--A--Q---K-V---G-...--G.-T-NIRFTN
A_RW.SF1703      -K----------------N-T..-E-- ----.-------...---...----..-----...-----...-----Y- ----AD-----G--N--KS...--S.---....----.S---------....-------------D
A_RW.W2RW008           -----V---------..---- ----.-----T-...I--...----..S-H--...-----...------- ----S---EA-R--VE---G...H-G.---....M--T.---
A_RW.W2RW009           -------------NET..-Q-- -S--.-------...-H-...----..-----...-V--...-----Y- T-NG-K--R---K--EK-SH...--E.-I-...T---K.---
A_RW.W2RW016           -------------D-A..-N-T -I--.-------...-P-...---R..-----...-----...N------ -----------EK--K--GE...H-R.---....-----.-
A_RW.W2RW020           -------------D-A..---- ----.------G...---...----..-----...-G---...------- ----AK--D--RG--KK--E...H-K.---....---E.K--
A_RW.W2RW021           -------------D-A..---- ----.-------...---...----..-----...-----...------- ----AK--D--RG--KK--E...H-K.---....---E.K--
A_RW.W2RW024           -L-----I-L---R--..---- ----.-----T-...---...----..T----...-----...----R-Y- TIN-----N------N--G-...H-G.KNK...T---T.K
A_RW.W2RW025           -------N--------..---- -V--.-----T-...---...----..T----...-E---...----R-F- T-NES---E--R------G-...H-E.---....-K-T
A_RW.W2RW026           -------------S-..---- ----.------R-...I--...----..-----...-----...---K--- ----S---N-----V---S-...--E.---....---N.S--
A_RW2.1613Wcons  -K-------D---------D-A..-Q-- ----.G------...---...----..-----...-----...---K--- T-NG----N--HK-----E-...H-R.---....-N-T.K-----
A_RW2.1613con    IK---K-L-D---------TEA..---- -S--.-------...-H-...----..-----...DA---...N------ ----A---N---K--Q---E...--V.---....---N.S---------
A_RW2.1701cons   -R-----------S----FX--..-R-- ----.-----T-...M--...----..-----...-G---...---K--- --T-S---R------K---E...--...-R-...---N.S---------
A_RW2.1831Wcons  ---------D---------KENV..T-T -I--.G------...I--...----..----S...-----...N------ ----AD-----XE--I--S-...--G.---....---T.------
A_RW2.1831cons   ---------D---------KEN..-T-T -I--.G------...I--...----..T----...-----...------- ----AD-----IE-----S-...--K.---....-----.K-----
A_RW2.561Acons   IK-------D---------D-A..-N-T -I--.-------...-H-...----..-----...-----...N----N- TI--AD--E--RK--I---T...--E.-T-...T---N.K-----
A_RW2.561Wcons   -K-------D---------D-A..-N-T -I--.-------...---...----..--F--...-G---...------- ----AD--E--XK--K---T...--N.-T-...T---N.K-----
A_RW2.R2235      ------------------T--..---- ----.-----T-...IH-...----..-----...-----...---K-Y- E-NK---TEA-K--SG--T-...--V.---....-K-D.KA--
A_RW2.R2235W     ----------------A--T--..---- ----.-----T-...IH-...----..-----...-----...---K-Y- E-NK---TEA-K--SG--T-...--V.---....-K-D.KA------
A_TZ.TAN14-2     -R-----------N----FAE-..---- -I--.-------...-H-...----..-----...-----...------- ----AP-REA
A_TZ.TAN15CON    -K-----------N-----TE-..-E-- -I--.-------...-H-...----..TS---...-----...------- -----K--T---K-VK--KQ...--G.---....---T.-
A_TZ.TAN8-1      -------------N----FI--..---- ----.-------...-H-...----..-----...-----...------- ----A---------VQ---T...H-G.---....-T--.R
A_TZ.TAN9CON     -R-----------N----FSR-..---- ----.------R-...-H-...----..-----...-----...---R--- ----A---T--KE--D---I...--E.---....-----.-
A_TZ2.016        -----------V-N----FTE..P---- -S--.-------...---...----..-----...-----...------- ----AR------K-VN---T...--G.---....-----.K-A-------....-------
A_TZ2.017        -----------V-N-----TE..P-Q-- ----.S------...---...----..-----...-----...------- ----AG------E--K--ST...H-E.---....----.--T-------....--------S
A_UG.1033         ------------N----FTE-..-T-- ----.-------...IHL...---R..-----...-----...------- ----AD-----KE-VK-
A_UG.964         -T--A---------N----FAR-..-R-- ----.-----E-...---...----..-----...-----...------- ---KE---N---E--K--Q-...ELG.-NK...T----.-----------....--------GE-FYCN
A_UG.U455        IR-----F------------N-..---- -S--.Y-.---NIRRYS-...-S--..---V-...-K---...------- ----RD--R-I----E--K-...K-N.---....----.S-----I---....--------------
A_UG.UG06        -K------SD---------T--..-T-- -----.YKKV-RR...IH-...---R..S--.-...SNL-...-----Y- ---GSQ--------V-K--E...-W..-T-....-N-S.KP--------....------V-------
A_UG1.W2UG029      --SD---------T--..-T-- ----.----ISRRR..IH-...-R--..------...-GMT-...N----Y- ---GSD--T-----V----R...HWQ.-T-....-V-K.S--
A_UG1.W2UG031      --S----------HT-..-N-S -I--.-------...---...----..-----...-----...N------ T-K-S------T--VKE---...-W..---....---T.---
A_UG1.W2UG037       -----V-N----NES..-T-- ----.-----R-...---...----..T----...-----...------- ---GSQ-----H--VE----...-W.N-N-....---N.S--
A_UG2.111        -------L-----N-----T--..---- ----.-----T-...-H-...----..---R...-----...------- ---KV---------VS----.
A_UG2.115        -------------N-----TE-..-I-- ----.-------...IH-...----..---R...------...N------ ---KA.--------VS---T...--...-T-....----.--T---P-
A_UG2.116        -------------N-L--FNTS..-Q-- ----.-------...-H-...----..-----...-E---...------- ---G-A------E--K---E..SLGN.---....---D.K-----P--
A_UG2.117        I------------N-L--FNET..-Q-- ----.-------...---...----..T----...-----...N------ -I-GAK--------KK-GD...LL.---...T---N.T---
A_UG2.118        -R---K-------N----FKT-..---- ----.------G...-H-...----..-----...-----...N-R--- E-N-----R---E--K---T...S-G.---....----.--T---P--
A_UG2.119        II-----L-----N-----KN-..-I-- ----.-----Q-...-H-...---K..-----...-G-V-...-----Y- ----AA------G--L----...H-E.---....---T.------P--
A_UG2.124           -SD---------FNES..-T-- ----.-------...-H-...----..-----...-A---...S------ -----K-REA--K--D--E-...H-K.---....
A_UG2.72         -------------N----FT--..---- ----.-----R-...-H-...----..-----...-----...------- ----K---N---E-VK---T...--G
A_UG2.78         -K-K----SD----L----T--..---- -----.G----R-...---...----..T----...-----...------- ----S--KE---K-VK---T...HW..---....---T.---
A_UG2.82         -V-----F-----N----FT--..-N-T -I--.-----T-...---...----..-----...-----...---K-Y- E-NK----R---E-V----T...E-R.---....----.--T---P--
A_UG2.84         I------------N----FTE-..---- ----.-------...IG-...----..T---A...DN---...------- ---G-K------E--K---E..HFGR.---....----.------P--
A_UG4.UG273      -K-K----SD----L----TT-..---- -----.G----T-...---...----..-----...-----...------- ----S--KE---K-VK---T...HW..---....---T.-----------....--------------
A_UG4.UG275      -K------------N-L- -TT-..-T-- ----.-----T-...---...----..S----...-----...--K---- -I--A---E---K-VS---T...H-G.---....--G.------I---....--------R-----D
A_ZR.Z321        -R-----F-D---I---------..-N-T -M--.-------...IS-...---R..--F--...-----...------- --------D--SK--A-----..HFVN.TS-...D----.------V---....--------------
A_ZR1.6557.sh         ----S-----------N-..-E-- -----.S--R-Q-...---...----..-----...-----...---R--- T-N--K--N--D--KA--E-...--N.-T
A_ZR1.6559            ---------------DE-..---- ----.----SQG...-H-...----..V---R...DR---...---K-Y- E-N-AV--N---R--A--G...S--G.---....-N-T.R---------....------A-------
A_ZR1.6563            ---F-D-G--------E-..-N-S ----.-------...---...----..T----...-----...-----Y- TIN-------------R--E...--T.TN-..TT-----.-H----
A_ZR1.6569       -A-------------------Q..-R-- -I--.-----RG...IHM...----..IL---...-S---...------- ---KKD--EA-HK-------...NH-N.---....---N.S---------
A_ZR1.6571.sh         -A---D---N----FNES..-Q-- ----.-------...I--...----..V--TN...-----...------- -I---A-YA---K-VK----...H-P.-
A_ZR1.6649            ----------------T-..-P-- ----.-------...-H-...----..-----...-----...N------ ----A--S----D------T...H-N.---....---N.RP---
A_ZR1.6653            ---F------L----DS-..---- ----.-------...I--...----..-----...-----...N------ ----AA--N---K---K---...--G.-R-....---N.S---------....----
A_ZR1.6655.sh         ---------------NES..-S-- ----.-------...I--...----..V----...-
A_ZR1.6657            ---------------F-N-..-Q-- ----.-------...---...----..T----...-----...------- ----E---E--HK-VE--GT...--R.-T-...N-T--.S---------....--------------
A_ZR1.6663            ---F-----------N-T..-Q-- ----.-----T-...---...----..-----...-----...---R--- ----AA--N--HK--VE-HT...L-A.-T-....VV-T.S-V--
A_??.5393        -Q-----------------FT-A..---- ----.-KT---G...---...----..-W--R...-NM--...---K-Y- --------Q---K-----G-...H-..---....-T-T.K---------....--------------
A_??.BLR10A      -I---------V-----H--T-..-T-- ----.-----T-...I--...----..-----...---T-...K-T--C- I-NG-V-ST---K--EK--G...--D.---....----
```

```
                                                      ______________________ V3 LOOP ______________________
                      ^^^          ^^^     ^ *^    ^^^                                             * ^^^    ^^^                 ^^^       ^ ^^                 *      *
B_CONSENSUS-95  VVIRSENFTDNAKTIIVQLNES..VEIN CTRP.NNNTRKS...IHI...GPGR..AFYTT...GEIIG...DIRQAHC NLSRAKWNNTLKQIVIKLRE...QFG.NKT....IVFN.QSSGGDPEIV....MHSFNCGGEFFYCNT
B_AU1.MRC1      ---------N---------KKP..---- ----.------D..---..----..-IFR-...-----...------- -V-GT---D-----------...--K.---....----.R---------....--------------
B_AU1.MRC2      ---------N---------K-P..---S ----.--------...-P-...----..--W--...-----...N------ KVNET--KD--R--AE----...--K.---....-I--.---------E....--------K-----S
B_AU1.MRC3      -----D-------------KDP..-N-T -V--.-----R-...---..----..---A-...-D---...------- ----ED-HKA-E--AG----...--N.---....----.R-----L-V-....V-T------L-----
B_BE.SIMI84     --------------------K-..-V-- ----.--SK--T...L-M...--K-..---A-...-D-G-YTGY------ -I------D------K--K-...---R---....---K.-----------....---------------S
B_BR1.10553         -----N---I-----KD-..-T-- ----.--------...-Q-...----..-----...-----...------- .----D--K--E---KN-G-...----.H--....-A--.R----------....T-------
B_BR1.10561         -----K-------------..-V-- ----.------G...--M...---S..T--A-...-----...------- ----T--E---R-----I--...--R.---....----.------S---....-------------
B_BR1.10565         ----------------Q-..---- ----.S-----G...--L...-F--..-L-A-...---T-...------- ---ST--EQ----VAR--HD...---.---....-F--.-----------....---V
B_BR1.10575         ----------------K-..-V-- ----.--------...LSL...----..-----...-D-V-...------- -----E-T---N---KE---...---.---....----.H----------....R-----R------
B_BR1.10581         -----N-------H--D-..---- ----.S-----G...---...----..T-FA-...-----...------- -I-E-------R-VA-----...---..-G-....-H--.-----------....--T---A------
B_BR1.10591         -------V-V------D-..L--- ----.----S-G...--M...---G..-----...-R---...---R-Y- -I-----------AR--G-...--V.---....----.-----------....T---------S--
B_BR1.10593         ---I-----------KDP..---- ----.--------...-S-...----..---A-...-D---...------- -I------D--Q---K---Q...--K.---....----.------Q-V-....-------------
B_BR1.7930          -----N-------H----..---- ----.------G...--M...-W--..T--A-...-----...A---P-- -I--E---K--E--AK--Q-...---..-I-...T-H--.R-A------R....--T----------
B_BR1.7932          -----N-V-----------..---- ----.--------...--M...-W--..---A-...-D---...------- -------EK-----AG--AK...LY.N---....-I--.-----------....T------------
B_BR1.7934          -----N-------------..-V-- ----.--------...--V...-W--..SLF--...-----...N--L--- ----I--E---N--A---K-...--V.---....----.R----------....T------------
B_BR1.7936            L-N-----------T..---- ----.--------...--M...-W--..---A-...-----...---E--- -I-------------K--K-...---.---....-I--.H------
B_BR1.7940          ---------S---H--V-..---- ----.------R...-Y-...----..-V---...-Q---...---R--- -I---------R---------...--K.-R-....---K.P----------....T-----------
B_BR1.7942          ----SN----------KT..-N-D -E--.--------...-N-...----..-----...-D---...------- -V-KND-L-A----A---G-...---.---....----.-----------....------R----F-
B_BR1.7946          -----N-----------T..-K-- ----.S-------...--L...-L--..-----...-D---...------- -I---E-----R--AQ---K...--E.---....----.H----------....--T-----
B_BR1.8615          -------T-I-----KDH..---- ----.H----R-...-T-...----..-----...-D---...------- -V--VQ-----A---Q----...--D.---....----.-----------....--T-
B_BR1.8621          -------T---------T..---- ----.S-------...--L...-W--..---A-...-----...------- ---T---E---E--AK--ED...---.---....----.-----------....-------------
B_BR1.8623          -----N----------TT..-K-- ---L.-----T-...---..---Q..---A-...-D---...------- -----D--D------T--G-...----.SN-...ITI-T.S------
B_BR1.8625          Y----N-------------..-A-- ----.--------...---..---S..---A-...-D---...------- -----E-D---T-V-K--A-...---.---....-K--.E----------....----I
B_BR1.8629          ---L-----------T-..I--- ----.--------...--M...-W--..T--A-...-----...------- -I-TE---T--E--AK--VD...---.---....-I--.---------VI....--T----------
B_BR1.8633          YG-L-G-T-----K-KDN..---- ----.------G...---...----..---A-...--T-...------- -V--T------A---K----...--P.---....----.--AE------....------
B_BR1.8635          ---I-----------KT..---- -L--.S-------...--M...-W--..---A-...-----...------- ---K---EE--N-VA-----...--EN-NK...IT--K.-----------....-------------
B_BR2.W2BR003         -I---K---------T..---- ----.--------...--M...-W--..---A-...-----...N------ -----------R------K-...H-..---....-I--.---
B_BR2.W2BR004         -L---T-I---------..-V-- S---.G-----G...-P-...---G..S--A-...ER---...------- X---T--E-----VAD----...--K.---....----.---
B_BR2.W2BR014         --MN----------N-..---- -I--.-----R-...-I-...----..---A-...-D---...-----Y- T-NSE----------T----...K--.---....----.HA-
B_BR2.W2BR017         ---N----------K-A..-Q-- ----.--------...---...----..---A-...-D---...------- -I---Q---------E---K...--D.---....-A-K.P--
B_BR2.W2BR018         ---N-----------T..-Q-- ----.----X---...---...---S..---A-...-D---...------- -----Q---------K----...--K.-S-....-A-K.P--
B_BR2.W2BR019         ---N-----------T..---- ----.G---R-...--M...-W--..---A-...-D---...------- -------E------ATX--G...--P.---....----.---
B_BR2.W2BR020         -L-N---------KDP..-D-- ----.--------...---...----..---A-...-D---...------- -----Q--D--SK--T----...--E.---....-K-Q.PP-
B_BR2.W2BR021         -------I-------T..---- ----.--------...--M...-W--..---A-...-----...N------ ---I---S------ER--KA...---..-N-....----.H--
B_BR2.W2BR023         ---N-------------..---- ----.--------...---...---K..---A-...-----...N----Y- -I--TA-RK--REVSR--A-...H-P.---....-I--.---
B_BR2.W2BR024         -----Y---------T..I--- ----.--------...--L...-W--..---A-...---V-...------- ---KDI-EK---R-A---G-...--K.---....-A--.K--
B_BR2.W2BR026         ---S-T---------P..---- ----.--------...-T-...----..---A-...-----...------- I-NKT--ES--E---KN-Q-...---.---....----.--T
B_BR2.W2BR028         -L-N----------T-..I--- ----.--------...--M...-W--..T--A-...-----...------- -I-SN--RT--G--AK--.-....X..-.X....-I.Q...V
B_BR2.W2BR030         ---------------P..-Q-- ----.S-----G...---...----..---A-...-D---...------- -I---Q---------G----...---.---....----.---
B_BR3.RJ12      I------L-N----------K-..-V-- ----.S-------...---...-W--..-I-A-...-A---...------- ----K--E---E--AK----...---.---....----.----------
B_BR3.RJ14      --V------N---I---H---T..---- ----.--------...--V...-W--..-L---...-----...N------ -I-GT--G----W-AA--K-...---.---....-K-Q.PPQEET
B_BR3.RJ19      G-----D--N---I---------..---- ----.--------...-QY...-T-G.G--A-...---V-...------- -I-KE---------AT--K-...---.---....----.-----
B_BR3.RJ379     -----D-------I---------..-V-- ----.G-------...---...----..-----...-D---...------- --------D--R--A-----...---..-G-..ET-I--.-----E-
B_BR3.RJ478     ---------N----------K-..---- ----.--Q----...--M...-W--..--H-N...-----...N------ ---GIE-----T--AE----...---.---....-I-T.-P----
B_BR3.RJ483     I--------N---V---------..---- ----.------G...--M...-L--..---A-...-G-V-...------- -I-----K---R--AK--K-...--N.---....----.-----------....-------------
B_BR3.RJ484     --L--------V-----------..---- ----.S-----G...---...-W--..---A-...---T-...---K-Y- -I--T--E---R--A---E-...--K.---....-I--.-----------....T-----------D
B_BR3.RJ49      -------I-N-------------..-V-- -S--.--------...--M...-W--..-----...-----...------- -V--ED-K---EK-AR----...H-K.E-P....-----.KT----
B_BR3.RJ54      -------I---------------..-V-- ----.--------...---...----..-----...-----...------- -----Q------L--T----...---.---....-A--.S------
B_BR3.RJ59      ---------N-------------..---- ----.G-------...--L...-W--..---A-...-A---...------- T-NKT--EK--E---------...--V.---....-T--.------
B_BR3.RJ62      --V------N---I---------..---- ----.S-------...--L...-W--..---A-...---V-...---E--- ----T----------A---K-...---.---....-E-K.-N----
B_BR3.RJ623     --V----------I-------T..---- ----.S----R-...--L...---G..-----...-----...N--K-F- T-NGT--E-------AE----...---.---....-A-K.-N-----
B_BR3.RJ626     ---------N-------------..IA-- ----.--------...-R-...---S..---A-...-D---...------- ----T--EH--T--GR--Q-...--N.---....-I--.-----------....-------------
B_BR3.RJ636      G-------N---I------Q-..-V-- ----.--------...-P-...A--S..-WFA-...-----...------- ---KVH-E-------AK----...--E.---....-Q--.---------
B_BR3.RJ64      ---------N-------------..---- ----.--------...--L...-W--..-----...-Q---...E------ -------EK-----A-----...---.---....VI--.-----G-RNC....N
B_BR3.RJI02     I-V------N---I---------..---- ----.--------...--V...-V--..-I-A-...-----...------- -I-ETR-K---EK-AK--K-...-VV.--K....-S-Q.-A------
B_BR3.SPB1      I--------N-------H----..-V-- ----.S-------...--M...-W--..---A-...-----...---R--- -----A--S--R--R
B_BR3.SPB2      I----K-----T-I-------T..---- ----.--------...--M...-W--..-----...-D---...------- -I-
B_BR3.SPB4      -G------------M-----TP..---Y ----.------R...KS-...----..-----...--V--...------- ----KR--D--I---TE---...H-K.---....----.-----------....I-C----R
```

V3 Region Alignments

```
                                                       ______________________ V3 LOOP ______________________
                    ^^^          ^^^      ^ *^    ^^^                                         * ^^^    ^^^                  ^^^       ^ ^^                 *      *
B_CONSENSUS-95  VVIRSENFTDNAKTIIVQLNES..VEIN CTRP.NNNTRKS...IHI...GPGR..AFYTT...GEIIG...DIRQAHC NLSRAKWNNTLKQIVIKLRE...QFG.NKT....IVFN.QSSGGDPEIV....MHSFNCGGEFFYCNT
B_HT2.H5996      --------N---I--------..---- ----.------G...---...---G..---A-...-G--...------- -I---------NK--K----...TY-.D--....---K.P-----A---....T-----Q-------
B_HT2.H5998      --------N---I--------..-V-- ---L.-------...-N-...----..---A-...RD--...------- -----E ----EK--KE--K..IEQFE---...----.-----------....--T---------D
B_HT2.H6000      -----------------KKP..-Q-- ----.-------...---...----.S----...-D--...------- -I--T---D----V-D----...--K.---....----.H---------....-------------
B_HT2.H6002      -----------------K-P..-Q-- ----.-------...---...----..-----...-D--...------- -I--T------R-V-D--K-...--K.---....----.H-----------...-Y-----------
B_HT2.H6004      -L------E-T-----H----..-V-- ----.-D-----...-PM...---K..---A-...-D---...N------ ---K-D-KK--HKV-K--G-...--K.H-Q....----.H----------.....---------I
B_HT2.H6006      I---------------H----..---- ----.-------...---...----..-----...-S--...------- -I-GT---A--EK--K--K-...--E.---....----.-------------...-------------
B_HT2.H6008      I-------N-----------T..I--- ----.------G...-T-...----..---A-...EK--...---R-Y- I-NKT--E------AK--A-...----.-E-MNKT-I-K.P-A--------....A------------
B_HT2.H6010      --S---L---------H----..-Q-H -I--.-------...-P-...----..---A-...-D---...---K--- -I-ETD-----IH--T---Q...--..---....-I--.S-----S---....T-----R
B_HT2.H6012         ------------I---A-..---- ----.-------...-P-...----..---A-...-D---...-----Y- --N-TQ-DG-----T----...---R--....-I--.S-T-------....-------------
B_HT2.H6014         -----N---I---------..-V-- ----.-D-----...---...----..-----...-Q---...N------ -----R-E-A-GK-AE----...---N--....----.-----------....-------------
B_HT2.H6016         -----N---N---------..-Q-- ----.-------...--M...---S..---A-...-D---...N------ T-N-T---S------K----...---.---....----.-----------....L-N---------
B_HT2.H6018         ---------------P..-V-- ----.-------...-P-...----..--F--...-D--...------- -I--EE--K--QK--T--G-...--R.---....----.-----------....K-----------
B_HT2.H6020         -----S-QN------KAP..---- ----.-----R-...---...----..---A-...-D--...------- -I--VQ--S--Q-LAA--G-...---
B_HT2.H6022         ---------------KNP..---- ----.S-----G...---...---G..-----...-----...------- -I--V------L-T------...---.---....-N-T.-----------....-------------
B_HT2.H6024         -G-L-S-----------T..---- ----.S------...---...----..---A-...-D---...------- -I------G--GR-AR-----...-V.-R....----.-----------....-------------
B_IN.IND9        ----------D--S------A-..---- ----.-------...---...----..---A-...-Q---...------- -I------D--R--S---E-...----.---....----.S----------....--------------
B_IN1.20016      -------------------K-..---- ----.K-EIKRRR..-K-...----..--VA-...-T.V-...-T---Q- --------E------------...--
B_IN1.20023      -----D-------------K-..---- ----.--SIKRRR..---...----..--FA-...NT.V-...-T---Q- --------E---------
B_IN1.30005      -----D-------------K-..---- ----.D-EI-R-R..LQV...----..--VAA...-T.A-...-T---Q- -------IE---L-------.
B_IN1.30008      -----D-------------K--..---- ----.G----R-...---...----..--FA-...-D-T-...------- ----T---E-----------...--Q
B_IT.Sala1       ---------N---S--------..---- ----.S------...--M...---K..T----...-D---...------- S--K----------A-----...---.E--....----.R----------....--------------
B_IT.Sala2       -----A--S--T-S---H----..-K-- ----.-------...-P-...----..Q----...-Q---...------- -----Q--K--ER-------...---.K--....----.R----------....--------------
B_IT1.115        I------IS--T-------K--..---- -A--.S----R-...---...----..-----...---T-...------- -I--E------R--AG----...--E.---....----.--G-
B_IT1.130        --F--------T-----------..-A-- ----.-----R-...-S-...----..---A-...-D---...------- -I-E----------K--G-...--V.---....----.H----
B_IT1.136        --X------N--RS------Q-..-Q-- ----.-----R-...---...---K..-----...-----...N------ -------DD---R-AD----...-YE.---....----.-----
B_IT1.145        ---------N----------T-..---- ----.-------...-P-...----..-----...-----...------- -I------D------S----...--Q.---....----.-----
B_IT1.190        -----K--SH-HES---------..-A-- ----.-------...-P-...----..-----...-D---...------- --------D--R--A-----...---.---....----.-----
B_IT1.193        -----KXXX--T---------T..---- ----.------G...---...----..---A-...-D---...N------ --------D---------K-...---.---....-A-K.-PA--
B_IT1.196        --------SN---V------K-..I--- ----.-------...VTV...----..VW---...-----...------- ---GXQ---------V----...---K--....-I-K.-P---
B_IT1.199        -I---K-----T-V---------..---- ----.G------...---...----..---A-...-----...---K--- -I------D------T----...-YE.---....-I-K.-----
B_IT1.201        -----A-L-----I--XH----..XV-- ----.S----XG...---...----..--F--...--T--...---K-Y- -V-XX---DX--X-A---G-...---.K--....-X--.-----
B_IT1.204        -----A-------S-------T..---- ----.-------...-P-...----..-----...-----...N------ -------EG---K-------...---.K--....-N-T.-----
B_JP.GUNA        I--------N-------------..IV-- ----.-------...-T-...-S--..-HAI...EK---...N------ -I-KEN---------T----...---.---....----.K---------....--T----------
B_JP.JH32        -------------------K-P..-V-- ----.SKT--RR...---...----..-----...KQ-A-...-L----- -IN--R--A------G---K...--V.---....----.R---------....-------------
B_NL.114C        -------------------KT..---- ----.-------...-P-...----..-W---...-D---...------- ---ITQ--K--E--AM----...--K.---....----.-
B_NL.168         I--------N---I------V-..-D-- ----.---I--R...---...----..-----...-Q---...N------ -I-KTR-----N--AK----...K--.---....----.----------....--------------
B_NL.X1con       --V------N-------------..I--- ----.--Y---G...-R-...----..-VIA-...EK---...------- ----TE-EK-----AN----.QS--K.---....-I--.-
B_NL1.A1         I----D-I-----------K-A..-Q-- ----.-------...---...---K..---A-...-----...------- ----VD-ED-----AE----...--R.---....----.-
B_NL1.A10        -----A-----------------..---- ----.-------...---...---S..-----...-D---...------- -I--T------Q---K-----...--N.---....----.-
B_NL1.A13        -I---A---N----------K-..---- ----.------R...VTM...----..VW---...-----...N-K---- -----E-E-----VAE---G...---N---....-A--.-
B_NL1.A2         ---------N-----M---K-P..-V-- ----.------G...--L...---G..T--A-...-----...------- ---ST---D--R---T-----...--V.---....----.R
B_NL1.A3         -A---S-----------------..---- -I--.-------...-N-...----..-----...-----...------- -I-G----S---M--E----...---.---....-N--.-
B_NL1.A4         I-----------------K--..---- ----.-----RG...---...-L--..R----...RK---...------- -I---------R---D----...---.---....-I--.-
B_NL1.A5         ---------N-------------..-V-- ----.H------...---...----..-----...-----...------- ---SV-------R---R--K-...---.---....----.-
B_NL1.A7         ---------N-----------T..---- ----.G----R-...-P-...---K..--F--...-----...------- -I-KG----M---V-------...---E--K...----.-
B_NL1.A9         -----A-----------------..---- ----.-------...---...-L--..-----...-D---...------- -I-GT------Q---K-----...--N.---....----.-
B_NL2.Wolfs-h1   ----D-I-----------KKA..-Q-I ----.-------...-P-...----..---A-...-----...------- ----VN-EI--H--AE----...--R.---....-A--.-
B_NL2.Wolfs-h495 ----D-------I----------..---- ----.-------...-P-...----..-----...-----...------- -I---------G---K-----...--R.---....-I--.-
B_NL3.NET1       -----------T--------K-..---- ----.-------...---...----..-----...-----...N------ -I------------AV----...--K.---....----.H
B_NL3.NET2       ---------N---------K-P..-R-- ----.-----R-...-G-...----..-I-A-...DR-V-...N------ ---K-E--R--E---K----...--R.---....----.-
B_NL3.NET3       -----D--S----------KDP..-N-T -I--.-------...-S-...----..---A-...-----...N------ ---GTE---A---VAE--SK...---.---....---K.-
B_NL3.NET4       -------L-------------V-..---- ----.------G...---...----..---A-...ER---...N------ -I-SE--------LV-T-----...K--N---....----.-
B_NL3.NET5       --------SN---------K-P..-K-- ----.-----RG...---...----..-V---...-K---...------- -I--T---D--Q-------K-...---.---....----.-
B_NL3.NET6       --M------N---S---------..---- ----.S----R-...---...----..-----...-Q-T-...N------ S-----------------K...--E.-R-....----.-
B_NL4.B63_84      ----D--------I------K-P..---- ----.-------...-Q-...----..-----...-----...N------ -I---------Q---------...---.---....----.--
B_NL4.BA1_84      ----A---N-----------K-P..---- ----.-------...---...----..-----...-D---...------- ---SS--YD--R-VAK----...---.---....----.--
```

V3 Region Alignments

```
                                                                    _______________________ V3 LOOP ______________________
                         ^^^            ^^^     ^ *^    ^^^                                          * ^^^    ^^^                  ^^^       ^ ^^                   *      *
B_CONSENSUS-95  VVIRSENFTDNAKTIIVQLNES..VEIN CTRP.NNNTRKS...IHI...GPGR..AFYTT...GEIIG...DIRQAHC NLSRAKWNNTLKQIVIKLRE...QFG.NKT....IVFN.QSSGGDPEIV....MHSFNCGGEFFYCNT
B_NL4.BA206_84  ----D--------------X--..-V-- ----.-------...---...----..-----...-----...------- --------D--E-V-X----...---.S--....----.--
B_NL4.BA65_85   ----D--------------KAP..-V-- ----.------G...---...----..-----...-----...------- ----TQ--DP-Q--A-----...--K.---....----.--
B_NL4.BN130_85  ----D-----------------..---- ----.-------...---...----..---A-...-----...------- ---K---E---E---K----...--K.-E-....T---.--
B_NL4.BN132_84  ----V--A------------G-..---- ----.------R...MTL...---K..V----...-----...------- -------D-----AT-----...--V.---....----.R-
B_NL4.BN142_84  --------N---I---------..---- -I--.-------...---...----..-----...-----...------- -I-K-R-----E---K----...---.---....----.--
B_NL4.BN152_84  ----D---N--N-------K--..---- ----.-------...---...----..-----...-----...------- ----E--D-A-E--DG----...--S.---....----.--
B_NL4.BN18_84   ----D--S------------A-..---- ----.-------...---...----..---A-...--V--...------- ----T--D-I-E--AKN---...--R.---....----.H-
B_NL4.BN29_85   ----D------R--------A-..-V-- ----.------G...---...----..-----...-D---...------- ----TE-D------AT----...---.---....-I--.--
B_NL4.BN90_85   I---------------------T.---- ----.-------...---...----..-----...-----...------- -I--E----M-Q--------...--E.-R-....---K.R-
B_NL4.BN93_85   --------N---I---------..---- -I--.-------...---...----..-----...-----...------- -I-K-----------K----...--V.---....----.--
B_NL4.H1091_90  ----------V---M-------..---- ----.-------...-P-...----..-----...-D---...N------ ----T--D---Q---Q----...---.---....----.--
B_NL4.H1112_90  M---------------------..IK-D -I--.-----R-...-P-...-L-S..-----...-----...------- -INST--D-------T--K-...---.---....----.--
B_NL4.H135_90   --------N-------------..-V-- ----.-------...--M...---K..T----...-D---...------- -I--E------N---K----...---.---....----.--
B_NL4.H153_82   ----------------------..IA-- ----.-------...---...----..-----...-Q---...-----Y- ----E--D---E-X-T--K-...---.---....-I--.-P
B_NL4.H164_90   -------S--------------..---- ----.-------...-P-...----..-----...-----...--S---- TIN-----D------G---D...---.--P....----.--
B_NL4.H169_85   XXXA---------------K-A..-R-- ----.-------...---...----..---A-...-D---...------- -I-K-E-----R---T--K-...---.---....----.R-
B_NL4.H17_82    --------N---I---------..---- ----.-------...---...----..---A-...-----...------- -I------D--E--AT----...--R.I-P....-A-X.--
B_NL4.H186_85   ----D---N----------KTP..-N-T -I--.G------...-P-...----..---A-...-D---...------- -----D------E-AK----...K--.---....-A--.--
B_NL4.H187_90   ----------G-----------..---- ----.------G...-R-...----..-IAA...TK---...------- -I-GE----------T----...---.---....----.R-
B_NL4.H188_90   --------NSG---------V-..---- ----.-------...-P-...----..-----...-D---...------- -I-----D-------M----...---.---....----.--
B_NL4.H1_85     ----D-I------------K-A..-Q-- ----.-------...---...---K..---A-...-----...------- ----VD-ED-----AE----...--R.---....----.--
B_NL4.H228_90   --------N-----------K-..I--- ----.------G...---...----..---A-...EA---...---K-Y- -V---A-----Q-L-T----...---.---....----.R-
B_NL4.H23_82    ----D---N---I------K--..---- ----.------G...---...---K..-----...-----...------- -------G----E-AR----...--E.---....----.--
B_NL4.H293_85   -F--D-IS-----------K-A..---- ----.-------...---...----..-----...-----...------- ----------------E--K...---.---....-I--.--
B_NL4.H29_82    ---A--L------------K--..---- ----.-------...-N-...----..-----...-GL--...------- -I--T---D-----AT--K-...--A.---....----.--
B_NL4.H339_82   ----D--------------K--..---- ----.-------...---...----..-----...-----...------- --------EA-----G---K...--E.---....---K.--
B_NL4.H35_80    ----D---N----------K--..---- ----.-------...---...----..-----...-----...------- -------DG--R---K----...---.---....-I--.--
B_NL4.H36_82    ----D---N----------K--..---- ----.-------...---...----..-----...-----...------- ----E--D-S-E--AR----...--R.---....----.--
B_NL4.H434_85   ----D---N-------------..I--- ----.-------...---...---G..---A-...-----...------- -I--G--V-------T--K-...-Y-.-I-....---I.S-
B_NL4.H43_90    -------S------------Q-..---- ----.-----RG...---...----..-----...-Q---...N------ -I-GE----------V----...----K--....---K.--
B_NL4.H450_85   I-------N-------------T.---- ----.------G...---...----..---A-...-D---...------- ---IT---------AE----...--V.---....----.H-
B_NL4.H457_85   ----D-----T-----------..-K-D -IS-.-------...---...----..-----...-----...------- ---K-Q--K--T--------...--E.---....----.--
B_NL4.H466_82   ----D--------------K--..---- ----.-------...---...----..-----...-D---...------- -I----------L-L-----...--R.---....----.--
B_NL4.H490_85   -L------N-V--------K-P..---- ----.-------...--L...---K..V----...-----...------- ---E---D--E---R-----...--E.S--....-I--.S-
B_NL4.H495_85   ----D-------I---------..---- ----.-------...-P-...----..-----...-----...------- -I--------G---E-----...--R.D--....-I--.R-
B_NL4.H552_90   ----A-L---------------..---- ----.-------...---...----..---A-...-----...------- -I-G------Q--AR-----...--E.---....----.--
B_NL4.H557_82   ----D-------I-------G-..---- ----.-------...---...----..-----...-----...N------ ---K---ED------T--G-...--E.---....----.--
B_NL4.H594_85   --------N-------H-----..---- ----.-------...---...----..---A-...-D---...------- -I------------R-----...---DTE-....-I-K.--
B_NL4.H724_81   ----D--------------K--..---- ----.-------...---...----..-----...-D---...------- -I---------N--M-----...--R.---....----.--
B_NL4.H748_90   ----K-----TE-------K--..---- ----.-------...-N-...----..---A-...-----...------- -----------N-V-T----...L-Q.V--....-A--.--
B_NL4.I1020_86  ----D-----------------..---- ----.-------...---...----..---A-...-----...------- -------E---E--------...--R.-E-...SV---.H-
B_NL4.I1025_89  ----A-------I---------T.---- ----.-----R-...-P-...----..---A-...-N---...------- -I-G--------RLAR----...--R.---....---Q.P-
B_NL4.I1035_87  ----A-----T-I---------..-Q-- ----.-------...-N-...----..-----...-----...--S---- --------D------K----...--K.---....----.--
B_NL4.I1057_87  ----------------------..-V-- ----.F------...-P-...----..-----...-D---...------- --------E-----------...--K.---....----.--
B_NL4.I1083_87  ----D-----------------T.---- ----.-----R-...---...----..-----...-G---...------- --------D------TY---...---.---....---I.--
B_NL4.I1164_90  --------N-----------G-..-V-- ----.-------...---...----..-----...-D---...------- --------E------T----...--K.-R-....----.--
B_NL4.I117_90   ----D-----------------..---- ----.-----IG...---...----..---A-...-----...------- -I---------Q--AM----...--V.---....----.--
B_NL4.I138_89   ----A--S-ST-----------T.---- ----.-------...-NT...----..-----...-D---...------- -----Q---------K--GD...--R.---....-A-T.--
B_NL4.I141_91   ----A-----T-S---------..-K-- ----.S-----G...-Q-...----..-----...-Q-T-...------- -------D------------...--K.---....----.P-
B_NL4.I142_87   ----D---N---I-M-------..---K ----.------G...---...----..---A-...-----...N------ -I-----E---R--M--G--...--K.---....----.--
B_NL4.I221_90   ----A-----T-S------K--..---- ----.-------...-T-...----..-----...-----...------- ---K-E-----R------Q-...--K.---....----.--
B_NL4.I257_91   ----A-------I---------..---- ----.-------...---...----..-----...-----...------- --------D------T----...---.D--....V---.R-
B_NL4.I3007_86  ----D-------I-------AP..---- ----.-------...---...----..-----...-----...------- --------D-----------...--K.-S-....----.--
B_NL4.I3008_88  --------N---N---------..---- ----.-------...---...----..---A-...-----...------- --------D---R-----G-...--K.---....---K.--
B_NL4.I3048_89  --T--H--N---V---------..---T ----.-------...---...----..-----...-----...N------ -I-GT---D---K--V--KD...---.---....----.H-
B_NL4.I3057_91  -------------------K--..IA-- ----.-----RG...---...----..-V---...-----...N------ -IT-KE--E--T-V-N----...--K.---....---T.--
```

```
                                                               _____________________ V3 LOOP ______________________
                       ^^^          ^^^     ^ *^    ^^^                                        * ^^^    ^^^                 ^^^       ^ ^^                 *      *
B_CONSENSUS-95   VVIRSENFTDNAKTIIVQLNES..VEIN CTRP.NNNTRKS...IHI...GPGR..AFYTT...GEIIG...DIRQAHC NLSRAKWNNTLKQIVIKLRE...QFG.NKT....IVFN.QSSGGDPEIV....MHSFNCGGEFFYCNT
B_NL4.I3064_90    ----A--S----I-------T..---- ----.-------...---...----..---A-...-D---...------- -I-----DD--RR-------...---.---....----.--
B_NL4.I3072_90    ----N--S--T-S--------..-V-- ----.-------...-N-...----..--F--...-K---...------- --------D-----------...--K.---....-A-S.P-
B_NL4.I3077_91    ----A---N---I-------P..---- ----.S-----X...---...----..---A-...-----...------- --------D--EN-------...--K.---....-A--.--
B_NL4.I3084_90    --------N---I--------..---- ----.----S-G...---...----..-----...-D---...------- ----T---D---RV-V----...--R.---....----.--
B_NL4.I3091_91    ----N-------I--------..---- ----.------G...---...----..---A-...-----...------- --------------A-----...---.-I-....----.--
B_NL4.I33_90      ----A---N---I--------..-V-- ----.G---SRG...---...----..-----...XK---...------- -I---Q--------------...--K.---....----.--
B_NL4.I39_91      ----A--S--T---------T..---- ----.-------...-N-...----..-----...-D---...------- -----Q---------X--G-...--R.---....-A-T.--
B_NL4.I4004_88    ------------S------V-..---- ----.-------...-PM...----..-----...-D---...N------ --------D--R--------...---.---....----.--
B_NL4.I4008_86    ----D-------I--------..-Q-- ----.H------...-P-...----..-----...-----...------- -I--EN-E----K-------...--K.T-K....----.--
B_NL4.I5020_89    ---TA----------------..---- ----.------G...---...----..-----...-----...N------ -----D--DA---V------...--D.---....----.--
B_NL4.I5032_87    ------------N------A-..---- ----.-------...---...A---..---A-...-----...------- ---IT---E-----------...--R.---....-I-K.H-
B_NL4.I5037_89    ------L-----I---H----..---- ----.-------...-X-...----..---A-...-----...------- -----T--A---R-----K-...---.-R-....----.--
B_NL4.I5038_89    ----N--S--T-S--------..-V-- ----.-------...-N-...----..-----...-----...------- --------D-----------...--K.---....-A-S.P-
B_NL4.I7025_91    --------N---I--------..---- ----.-------...-P-...----..-----...-Q---...------- -I-----TD---K-------...---.-NI....SI--.--
B_NL4.I70_88      ---A------T--------K-..I--- ----.------G...---...---K..---A-...-----...N----Y- ----SD--EA---V-N----...--E.---....----.--
B_NL4.I90_89      ----A--SN---V--------..---- ----.------G...---...---S..---A-...-----...------- -----Q--D---R------D...--E.---....----.H-
B_NL4.K999_86     ----D---N---I-M------..---- ----.-------...---...----..-----...-----...------- --------D---R-------...--E.---....----.--
B_NL5.pt10con    ---------N---------KDA..--V- ----.-------...---...----..--F--...-D---...------- -I------S------K--E-...--K.---....----.-
B_NL5.pt11con    ---------N---------K--..-.NI -I--.-----RG...---...----..---A-...-D---...---K-Y- I-NST------S-V-D----...--K.---....----.H
B_NL5.pt12con    ----------D-----------..-V-- ----.------G...-R-GHI----..-----...-K---...---K--- -I--KD--------------...---.---....-T--.H
B_NL5.pt13con    -----D---N---I-M------..-Q-- ----.------G...---...----..---A-...-----...N------ -I--T---D---K-------...--K.--K...T---T.-
B_NL5.pt14con    II---------------H--K-..II-- -I--.-----R-...---...----..---AR...-D---...------- K-N-TE-----R---T----...--E.---....---K.-
B_NL5.pt15con    -------I---T-----H----..-K-- -I--.------G...---...----..---A-...-K---...------- -I---------GK-----K-...K--.-T-...T----.-
B_NL5.pt16con    ---------N----------K-..-K-- ----.S----R-...---...----..---A-...---T-...------- -I-G-E-----E---K----...---.---...T---K.-
B_NL5.pt17con    ---------N----------K-..IV-- ----.-----R-...-NM...----..---A-...-Q---...------- ----VE---------T---D...---.---....----.-
B_NL5.pt19con    ---------N---I--------..---- ----.-----R-...---...---S..---A-...-D---...------- -I-KE--D------A-----...--E.---....-R--.-
B_NL5.pt1con     -----D----------------..---- ----.-------...---...---K..-----...-----...------- -------DY--Q--------...--K.---....----.-
B_NL5.pt2con     ---------N---------K-P..-K-H ----.H----R-...---...----..---A-...-D---...------- ---KTD--K--E---K----...--E.---....----.-
B_NL5.pt3con     -I---Q---N------------..IQ-- ----.-------...-N-...----..---A-...-A---...------- -----Q---------T--K-...---.---....----.-
B_NL5.pt4con     -----D--S-----------K-..-T-- ----.------G...-N-...----..-----...-----...------- --------D-----A-----...--K.---....----.-
B_NL5.pt5con     -----A---N---I-----K--..---- ----.-------...-P-...----..-L---...-----...------- -I--T------Q--AK----...--K.---....----.-
B_NL5.pt6con     -------I---I-----H--A-..---- ----.-------...-P-...----..-----...-----...------- -I-KE---E--G--------...--R.---....---K.-
B_NL5.pt7con     -----A--S----------K-P..-I-- ----.-------...-PM...---K..---A-...-D---...---K-Y- E-NGTQ-----Q---K----.IE--K.---....----.H
B_NL5.pt8con     -----S--------------K-..---H ----.------R...---...----..-V---...-----...------- -----A--D---E-------...---.---....----.-
B_NL5.pt9con     ---------N------------..---- ----.-------...-N-...----..---A-...-----...---E--- ----E--EK------T----...--K.---....----.H
B_NL6.16         -----D----------------..-V-- ----.------G...---...----..-V---...-R---...------- -I-GV---K--E---T----...--R.---....----.----------....---------------
B_NL6.320        -----A--SN-------H----..-A-- ----.------G...--M...---K..---A-...-Q---...------- ---------------S---K...---.---....---S.-PL--------....---------------
B_NL7.NE1058con  I----D--S------------T..---- ----.-------...-P-...----..-----...-----...------- -I--T------N--AK----...-L..---T...----.----------....T--------------
B_NL7.NE537con   -----A------R---------..-V-- ----.-----R-...-N-...----..--W--...-D---...------- ---I-----------T----...---N---...----.---------I....--H------------
B_NL8.1058       I----D--S------------T..---- ----.-------...-S-...----..-----...-----...------- -I--T------N--AK----...-L..---T...----.----------....T--------------
B_NL8.537        -----A------R---------..-V-- ----.-----R-...-N-...----..--W--...-D---...------- ---I-----------T----...---N---...----.---------I....--H------------
B_NL8.594        ---------N-------H----..---- ----.-------...---...----..---A-...-D---...------- -I-------------R---D..---DT--....-I-K.----------....---------------
B_NL8.672        -----D----------------..---- ----.-------...--M...---K..-----...-G---...-----Y- -I-----D--R---T--K-...K--.---....----.----------....--------------S
B_PR1.D2PR728          ---N---------KDP..IK-- ----.-------...---...----..---A-...-----...------- -----E--S--H--------...---.---....-A--.R----------....--------------T
B_PR1.D2PR729          ----------------..-M- ----.-------...---...----..-----...-D---...------- -----S-----GH-A-----...---.---....-A-K.P---------....-L-----------ST
B_PR1.D2PR732          ---N---I-----K--..---- ----.-------...---...----..-----...-D-V-...-----Y- -I------D----VA-----...--K.---....----.----------....----------
B_PR1.D2PR733          ----------------..---T ----.-------...---...----..---A-...-Q---...------- -----G-----EK-------...--R.---....----.P-----
B_PY.12837             -IS--------H--D-..-Q-T -I--.--Y---R...VSL...----..VW---...-----...---K-Y- -V--TQ--Q--ERVRAE-KSHFPNATAI-LN
B_PY.12838             -----T--------KT..IT-T -V--.------G...-R-...--R-..GTVY-...TR-V-...------- ----TE--D------K--G-...----.-TS...T-S--
B_PY.12839             --S----I-------T..---- ----.------G...--L...-W--..T-F--...EK---...N------ ----ER-K------I-----...--T.---....---K.-
B_PY.12841             -IS--------LF-K-..-L-- --T-.-------...-F-...----..--H-A...-R---...---K--- -I-EKQ-SDV-RKVKAE-GSHVPNTTTI
B_PY.12842             --S--T-N-----KAA..-V-- ----.------G...---...---K..-Y---...-D---...------- -I---E--D------VE--K...--E.--
B_PY.12845             K--N------------..-K-- ----.-------...---...----..---A-...---V-...------- SI---E--RA-RRVA---G-...--P.---....-N-
B_PY.12847             ----S----------A..-K-- ----.-------...--M...----..---A-...-D---...N------ -I--EE-----QK-----Q-...---.---
B_PY.3614p             -IS--------H-K--..---- ----.-----TR...-QM...----..VL-A-...-D---...---K--- -I---H-S-A-D-V-TN--A.
B_PY.3615              ---N---V---H----..---- ----.--KIMRR...---...----..--H--...RN-V-...N----Y- -I--V---A--HK-------.
```

V3 Region Alignments

V3 LOOP
B_CONSENSUS-95
B_US.JFL
B_US.JM
B_US.JRCSF
B_US.MA
B_US.MN
B_US.NY5CG
B_US.P896
B_US.RJS
B_US.SBA
B_US.SBB
B_US.SBC
B_US.SC
B_US.SF162
B_US.SF2
B_US.SF33
B_US.TN-1000
B_US.TN-1001
B_US.TN-1002
B_US.TN-1003
B_US.TN-1005
B_US.TN-1006
B_US.TN-1007
B_US.TN-1009
B_US.WM
B_US.YU
B_US1.HC01
B_US1.HC05
B_US1.HC06
B_US1.HC07
B_US1.HC08
B_US1.HC09
B_US1.HC10
B_US1.HC11
B_US1.HC12
B_US1.HC13
B_US1.HC16
B_US1.HC18
B_US1.HC19
B_US1.HC20
B_US1.HC21
B_US1.HC22
B_US1.HC23
B_US1.HC24
B_US1.HC25
B_US1.HC26
B_US1.HC27
B_US1.HC28
B_US1.HC29
B_US1.HC30
B_US1.HC31
B_US1.HC32
B_US1.HC33
B_US1.HC34
B_US1.HC35
B_US1.HC36
B_US1.HC37

V3 Region Alignments

```
                                                            V3 LOOP
                    ^^^         ^^^     ^ *^    ^^^                                       * ^^^    ^^^                ^^^      ^ ^^                 *      *
B_CONSENSUS-95  VVIRSENFTDNAKTIIVQLNES..VEIN CTRP.NNNTRKS...IHI...GPGR..AFYTT...GEIIG...DIRQAHC NLSRAKWNNTLKQIVIKLRE...QFG.NKT....IVFN.QSSGGDPEIV....MHSFNCGGEFFYCNT
B_US2.D2US664         -------I-V-H--K-..-V-- ----.-------..---..----.---A-....D---..------- ---K-D-EEA-N-TAR--K-...-L..-T-...T---.----------...----------I--S
B_US2.D2US665         -I---------H----..I--- ----.-------...-PM..----.------..--V--...N----Y- -I-K----D------T------..H-K.---..---.----------...-------------D-
B_US2.D2US704         ---N-----------T..-V-- ----.-------..---..----.--H--..--V--..------- -I-GE------EK--K--K-..--R.D--...---K.H----------...----------Y----
B_US2.D2US705         ---N---------K-A..-K-- ----.-------..-N-..----.---A-..-----..------- -I--TD-----GK--E---A..-..-CS..K---K.----------...---------------
B_US2.D2US706         ----------------..I--- ----.-------..-N-..----.-----..-----..------- ---KE--ET-----AE---K...-Y..---..---K.----------...--
B_US2.D2US714         --S----N-L------..-N-T -I--.-----R-...--M..----.---A-..-D---..------- -I--T---E--QR--K--G-..---.---..P---.--------
B_US2.D2US717         ---------------P..-KMD -I--.-----R-...-N-..----.-----..-D---..N------ -----N-T---QK-A---G-..--Q.---..----.---------T....-I-----------SS
B_US2.D2US724         --SN--A-------A-..---- ----.G-K-IR-...-SM..----.---.R-..-Q---..N----N- T-------E--R-V-K--SK..---.---...-A-K.-P---------...T-------------
B_US2.D2US725   ------------------K-..---- ----.-------..-P-..----.---A-..-D---..------- -I------------E--K-..---.---..----.---------
B_US2.D2US726         --------------KP..-V-- ----.-----R-..----..A---.--HA-..-N---..------- ----TQ-D---H---T---K..---AD--..----.----------...---------------
B_US2.D2US727         ---N------------..-A-- ----.S------..V--..----.-----..-----..------- -I-AT---------D--K-..---.---..---K.----------...---------------
B_US3.D2US711   ---------N---------K-A..---- ----.-------..--L..----.---A-..-----..------- T-NKT---EA-----M---G..---.---...-I--.K--------E...Q--------------
B_US3.D2US712   -----A-LS----I---------..--M- -I--.-------..---..----.-----..-D---..---K--- -I------------A---K-..--E.---..----.---------M...TLM-----------S
B_US3.D2US715   -----A-LS--------------..QM- ----.-------..---..----.-----..-----..------- ----T---E---R-------..-YE.---..----.----------...-L-----------S
B_US3.D2US716   I--------N---------K--..--M- ----.-------..---..----.-----..-Q---..------- -----S-----EN-T-----..---.---...---K.P----------...-L-L---------S
B_US4.ZhuPtA    II---A--S-----------A-..-A-- ----.-------..-P-..----.-----..-D---..---K--- -I---E---------G--G-..--E.-A-..T----.R----------...T-------------
B_US4.ZhuPtF    -----K-----T-I-------T..---- ----.S----R-...--M..-L--.-----..-D---..------- -I--E-------K--T-----..-YE.--I....-I-K.P----------...-------------
B_US4.ZhuPtL    I----A----------------..-I-- ----.------G..---..----.-----..-Q---..---K--- -I---Q--G--EK--R-----..-R.---..-----.H-------V-...-------------
B_US4.ZhuPtR    ---------N----------Q-..-V-- ----.-----R-...-P-..----.-----..-Q---..------- ---GIQ-----E---R-----..K--N---...-I-T.---------T...T-------------
B_US4.ZhuPtV    L------------N---------..---D -I--.---------...-TM..---K..V--V-...D---..-----Q- T-N-T------E--AK-----..--.---..----.----------...-------------D
B_US5.pt5-bl         ---L-N-----------T..-T-- ----.S-----R...-A-..----.---V---..EQ---..---R--- -I-K-R-----R-VA------..KL-.---..---K.N----------...----T-----P---
B_US5.pt6-bl         -----N---------K-A..-N-T -E--.-------..-N-..----.---A-..-D---..------- -I-GT------N---------..---.---..---Q.P----------...T----------F--
B_US6.I01       -------I-E-T-----H----..---- ----.------G..--L..--G.---A-..-----..------- ---ES-----I--V-E---K..H--.---..----
B_US6.I02       -----D---N---I---------..-V-- ----.-------..-T-..----.---A-..-D---..------- -I--TQ-------V-T---Q..---.---..-A--
B_US6.I03       ---------N-------------..---- ----.-----R-...-N-..----.---A-..-----..------- -I-GV----------K----..---N---..----
B_US6.I04       I----A--S----I---H-----..-A-- ----.-----RG..-N-..----.---A-..-----..------- -I---E--K--EK-------..---N---...---K
B_US6.I05       -----D--S----------KDP..-N-T -I--.-------..---..----.-----..-----..N------ ---I-A-----Q---G-----..---.---..----
B_US6.I06       -I---D--S-----------K-..I--- ----.-------..-P-..----..W---..-D---..N------ -I-G-E---A----------K..---.---..----
B_US6.I07       --V------N-----L---R-A..---- ----.-------..--L..--S..V--.-----..------Y- T-NGS---D------T-----..-K.---...-A-K
B_US7.306        ------L-N---------K-P..---- ----.-------..---..---K.-----..-----..------- -I------D--R---
B_US7.552             ---N---I---------..---- -I--.-------..-P-..----.--FA-..-D---..------- -I---N--------T
B_US7.565        ----D--S-----------K--..---- -I--.-------..-P-..----.-----..-D---..N------ -I-K----D-----
B_US7.657        ----D--S-----------K-..I--- ----.----X--..---..----.-----..-A---..------- ----VQ-----Q---
B_US8.R1           --D-L-----I---------..---- ----.-------..---..----.-----..-----..------- -------ED----V-R-----..-K.---..---K.----------...T------------S
B_US8.R2        -----D--------------K--..---- ----.-------..-N-..----.-----..-----..------- -------E---GE-AKE--R..--E.---..----.----------...---------------
B_US9.S1        -----D---N-------H-----..---- ----.-------..---..----.-----..-----..------- -I-GT----------K--G-..--R.---..----.---------
B_US9.S2        -----------------------..-V-- ----.-------..---..----..W---..-----..------- ---E-N--KA-E---K--K-..---.---..----.---------
B_US9.S3        -------------------KD-..---- ----.-------..---..----.---A-..-D---..------- -----Q-----G---A-----..-K.-S-...-S-K.P--------
B_US9.S4        -------I---------H-----..-K-- -I--.-------..---..----.-----..-----..------- -I-EE---K--G------K-..--K.---..----.---------
B_VE.VE1        II-------N----------D-..-V-- ----.-------..--M..----.---A-..-----..-----Y- DI--T--K---M--AK-----..---N-T-..----.----------...T------------S
B_VE.VE2        -----V--S----I---------..---- ----.-----QG..---..----..-I---..-RT--..------- -I-K---K-----VAK-----..---.---..----.R-----L--E...------R--------
B_VE.VE3        -I-----L---T----I------..---- ----.------G..-G-..----..V--A-..-G---..------- -I--V------G--AK-----..---.---..----.--.--------...---------------
B_VE.VE4        -----Q---N---I---------..---- ----.-----R-...--M..-F--..-L--N..------..N------ -I-GT--------V-K-----..---.---..----.------L--T...T----------F-S-
B_VE.VE5        ---------N-------------..-V-- -I--.------G..-Y-..----.------..EK---..N--K-Y- T-NKTI-----EK--N-----..-V.---...-I--.P--------E...---------------S
B_VE.VE6        I--------N----------Q-..I--- ----.------R..-TM..----..VL---..-Q---..--KK-Y- -I------D--RKVA------..--E---...-I-K.P----------...T-------------
B_VE.VE7        ---------N---------K-A..---- ----.-----RG..--L..-L--..R-----..D.---..-T----- -I-T---------V-ER---..---.---...-VF.H----------...-------------S-
B_VE.VE8        -----A--S---R----------..---- ----.--Y-A-R..-S-..----..---R-K..-R---..------- TINK----E--GK--T-----..-R.---..----.----------...---------------S
B_??.LIT11A     -----------V-----------..IQ-- ----.-------..---..---K.---A-..-D---..------- ----TQ---------T-----..-K.---..----
B_??.LIT17A     ---------N---------S--..---- ----.-------..---..----.-----..-----..------- ----T---D--R--AT--G-..--K.---..----
B_??.LIT18A     -----------V-----------..IQ-- -I--.Y------..-P-..----.---A-..-D---..------- ----TQ---------T-----..-K.---..----
B_??.LIT21A     ---------N-------------..---- ----.-------..-P-..----.-----..-D---..------- ----T--DE--R--AT-----..-K.---..----
B_??.LIT22A     I------I-N-----------T..-Q-- ----.-----R-..----..-W--..-V-A-..-G---..---R-Y- T--G----S---R-AT--G-..............H--
B_??.RUS3A      -------------------K--..---- ----.-------..---..---K.---A-..-G---..------- -I-ET------R---T--Q-..---.-N-...----
B_??.RUS4A      -L-------N-V-----------..---- ----.-------..--M..---G.---A-..-A---..------- -I-SE----------T-----..---.---...-A--
```

V3 Region Alignments

V3 LOOP

```
                                                                ______________________ V3 LOOP _____________________
                        ^^^          ^^^     ^ *^    ^^^                                          * ^^^    ^^^                  ^^^                                    *      *
D_CONSENSUS_95 IIIRSENLTNNAKIIIVQLNES..VTIN CTRP.YNNTRQR...THI...GPGQ..ALYTT....RIIG...DIRQAHC NISGAEWNKTLQQVAKKLGD...LL..NKT...TIIFK.PSSGGDPEIT....THSFNCGGEFFYCN
D_UG3.121      V-----------T---------.-A-- ----.N----S..---..----.------....K---...---R--- E--E-G---------R---H...--..-T-....----.--
D_UG3.122      -------------T---R---T..-K-- ----.A-----S...V-L...----.------....-V--...-----Y- S--E----R------V----...--..-Q-...--N--.------
D_UG3.70       -V--------------------.-I-- ----.---IKIQ..R-P-...-R--..--F--....--K-...IKG---- ---K-G-----------Q-.
D_UG3.71       -VV-------------H--K-..---- ----.N----S...I--..----.------....NV--...------- ---R---------TE--KY...-F..---....--NF.E
D_UG3.74       -------F-----------K-..---T ----.-T-K--G...--M...---R..----I....D-T-...-----Y- ---K-K-YN------N----...--..----...--N-E.------
D_UG3.79       -------------T---------.---- ----.-----S...---...----.------....N--...------- -V-K-G-G----------K-...-F..KAD...N-T--.------
D_UG3.83       -------I----RT-----K--..-K-T -V--.-S-Q-R-...-P-...-L--.------..RMDNMK...N-K--Y- ---VED------------G...--..-L-...-----.------
D_UG5.UG266    -V--------------------.-I-- ----.---IKIQR..-P-...-R--..--F--....-RK-...IKG---- ---REG-----------R-...-I..---...-----.----------....-------------
D_UG5.UG269    --------------------A-..---T ----.-TK--H-.........AQ-R..-WW--....G-T-...-----Y- ---K-K-QE-----------...--..---...Q---E.--------V-....K-----------D
D_UG5.UG274    ----------------------.---D -A--.-E-I---...-P-...-T--..-----QGRKK.--...K-G---- -V-K-G-E--------N
D_UG6.9801           --------------P..-I-- ----.-QKVHTY..RP--....G-R..-S---....--T-...N------ ---R----------N--RG.
D_UG6.9802           ---------------..---- ----.-QTI-H-...---...-S--..----P....--T-...Y------ -V-----------------.
D_UG6.9803           ---------------..---- ----.-K-I-H-...-P-...-S--..------....-AV...-----Y- ---K---------------.
D_UG7.42       V------I---V-T---L---T..-Q-- ----.-SKI---...-ST...-Q--..------....KGR-...I-G--Y- ----------HSSNILRRP...PS..QTK...INFLP.HPLE-T----....----I------S--
D_UG7.44       -------------T-----K-P..-N-T -I--.-----KS...---...----..-YF-S....E---...------- ---K-Q--N--T---N---G...-S..ST-...M---S.-------GKVH...Q--V----G----D
D_UG7.963      -----------------K--.S-P-- ----.------G...---...---R..-F---....E---...------- ---RVG-D---H---G--RA...F-..KTP...P---R.----------....-----------S--
D_UK1.CPHL4    -------I-D-V-N---------..-L-T -A--.S-----S...---...----..------...GQ---...------- -V-KKD-SN---R--I--KN...--..---....---N.T---------....--------------
D_US.AMK       -------I-----T---------..-S-- ----.-SKE-LK...-S-...-Q--..------..RVKVT-...------- -V-EIK---------A--KG...--..-ISS..--H-FK----------....-------------
D_ZR.ELI       V-----------N--AH-----..-K-T -A--.-Q------...-P-...-L--..S------...-SRS...I-G---- ---R-Q-S--------R---T...--..---...I-K--.-
D_ZR.JY1       -----------V-T---H-----..-E-- ----DNKI---S...-P-...-L--..------....--K-...-----Y- ---A-A-------------...--..-Q-...-----.-
D_ZR.MAL       -M-------D-T-N--------T..---- ----.G----RG...I-F...----..------....G-V-...---R-Y- T-NET--D-------V---S...--..---...K---N.S
D_ZR.NDK       -----------V-T------A-..IV-- ----.-KY----...-S-...-LR-..S---ITGKKKKT-...Y-G---- K--R-----A-----T---N...--..---...--T--.-
D_ZR.Z2Z6      ----------------------..-A-- ----.-R-I---...-S-...-L--..------....KTRS...I-G--Y- ---KN----------I---N...--..---...-----.-
D_ZR1.6555        -K-I---VNN---------..-A-- -I--.N-----S...--L...----..------....KV--...-----Y- ----T----A-----I---N...-F..---...--N-D.H---------....--
D_ZR1.6565.sh  ---------D-L-N---------..-I-- ----.N----KS...I--...----..-I---....DV--...-----Y- E-N-PK-------
D_ZR1.6651.sh     ---------T------AT..-A-- ----.N----KG...I--...----..----S...GD-V-...-------
D_ZR1.6659.sh    -------D---T---------..---- ----.N--V---...-P-...----..-F---...G
D_??.RUS14A    ----------------------..---- ----.---I---...-P-...-L--..------....-R-E...---R--- ---K-A-K---E------KV...-F..-R-...--D--
```

```
                                                  ______________________ V3 LOOP _____________________
                      ^^^          ^^^     ^ *^    ^^^                                     *   ^^^                      ^^^                                *      *
E_CONSENSUS_95 IIIRSENLTNNAKTIIVHLNKS..VEIN CTRP.SNNTRTS...ITI...GPGQ..VFYRT...GDIIG...DIRKAYC EINGTKWNEVLKQVTEKLKE...HFN.NKT....IIFQ.PPSGGDLEIT....MHHFNCRGEFFYCNT
E_CF1.1697     QNNQI$----------------..---- ----.FK-M---...AR-...----..---K-...-S-T-...------- ---------T-----K--R-...--K.---....----S
E_CF1.4002     ---------D------------..I--- ----.FKKV-I-...AR-...---R..--HT-...-N-N-...------- ---E---K-T-----R--R-...-L..-G-..TT-S-TT
E_CF1.4013     --------------------E-..-T-- ----.FKRV---...VR-...---R..--HK-...-A-N-...------- K-D------T--R--I--R-...--.---....-LYRP
E_CF1.4017     ---------------------A..---- ----.--KI---...VR-...----..---K-...-A-M-...-----F- ---------T--E-VK--RK...--.---....-H-RP
E_CF1.4031                  ---------..---- ----.-------...VR-...----..---K-...-----...---R--- ---------T-----K--R-...Y-.---....---KP
E_CF1.4039     VV-----I---------Q-VTP..-Q-- ----.---I---...VR-...----..---K-...-S-T-...-----F- --KEAQ--KT--E-IG--R-...--.---....----R
E_CF1.4069     --------AD----------E-..---- ----.FKK--I-...AR-...---R..--HK-...-A-L-...-----F- ---K-R--DT-NK-------...--..K--....-L--P
E_CF1.4084      N--------D-----------P..---- Y---.-------...VR-...----..---K-...-E---...-----F- K--E-----T--E--K--R-...--K.---....----P
E_CF2.4071TG2  ----------------------..---- ----.FKKM---...VR-...---R..---K-...-S-T-...------- ---------T-Q-IIR--E-...---..-NK..T-Q-KP-Y----------....---------------
E_CF2.ELO      ------------R---------..---- ----.YK-----...AR-...----..---K-...-S-T-...------- -----A-H-T---MVG--RK...--.---....----.-S----P---....--N------------
E_CF2.MBA      --------A------------P..---- ----.-------...VR-...----..---G-...-E---...---R-F- ---------T--------R-...--.---....-V--.-------------....--S------------
E_CM.CA10      ----------------------..---- ----.-TTI---...SR-...----..A--KI...EG-S-...N--A--- ---------TF---K--R-...Y--.---....-K--.-SP---P---....--S------------
E_TH.N764              --------------..---K ----.-------...---...----..-----...-----...N------ ---------T---AK--N-...--K.---....---H.-------VV-....---------------
E_TH.T8178         ------------------..---- ----.-------...-S-...----..-----...-----...------- ---------A-T----N---...--H.---....-L--.------------....T--------------
E_TH1.8639             --.-----------..---- ----.------R...---...----..-----...-----...------- ---------A----------...--.---....----.--------L-....--------------
E_TH1.8657          -----------------..---- ----.-------...---...----..I----...-----...------- ---------A----------...--.---....----.------------....--------------
E_TH1.8659            ---------------..---- ----.-------...---...----..-----...-----...------- --------------------...--.---....----.------------....--------------
E_TH1.8663          -----------------..---- ----.-------...---...----..-----...-----...------- --------K-----------...--.---....----.------------....--------------
E_TH1.8671       --------------------..---- ----.-------...---...----..-----...-----...N------ ---------A---AG-----...--.---....---H.-------VV-....------S-------
E_TH1.8673         ------------------..---- ----.-------...---...----..-----...---T-...N------ ---------A-R--------...--.---....----.------------....--------------
E_TH1.8683     ------E---------------..---- ----.-------...-P-...----..-----...-----...N------ ---------T----------...--.---....----.------------....--------------
E_TH1.TA10671          --------------..---- ----.-------...--M...----..-----...-----...---R--- ----K---KA----------...--I.--K....----.-------
E_TH1.TA7792           --------------..---- ----.-----P-...---...----..-----...-----...------- --------KA----------...--.---....--C-.------P-
E_TH1.TA7794           --------------..---- ----.-------...---...----..---K-...-----...N------ ---------A----------...--.---....----.------------....---------
E_TH1.TA8173           --------------..---- ----.-------...-P-...----..-----...-----...------- ---------A----------...--.---....----.------------....T---
E_TH1.TA8176           --------------..---- ----.-------...---...----..-----...-----...------- -V-------A----------...--.---....----.------------....----
E_TH2.TN235    ----------------------..---- ----.-------...-P-...----..A----...-----...------- -----------T--------...--.---....----.------------....------X-------
E_TH2.TN238    ----------------------..---- ----.-------...---...----..-----...-----...-T----- ---S-R---A-----G----...--.---....----.------------....--------------
E_TH2.TN239    ----------------------..---- ----.-------...---...----..-----...-----...N------ ---------A----------...--.---....----.------------....--------------
E_TH2.TN240    ------D---------------..---- ----.-------...---...----..-----...-----...N------ --------K-----------...--.---....----.------------....--------------
E_TH2.TN241    ----------------------..---- ----.-------...---...----..-----...-----...------- ---------------GE---...--.---....----.------------....--------------
E_TH2.TN242    ----------------------..---- ----.-------...---...----..-----...-----...--X---- ---R----A-----K-----...--.---....-L--.------------....--------------
E_TH2.TN243    ---------D------------..-G-- ----.----XP-...--X...----..-----...-----...---X--- --------R-----------...--.---....----.------------....--------------
E_TH2.TN244    ----------------------..-V-- ----.-------...---...----..-----...-----...------- -----E--KA----------...-----P....----.------------....--------------
E_TH3.W2TH001        ----------------..---- ----.-------...-N-...----..-----...-----...------- -----------R--------...--I.-R-....----.---
E_TH3.W2TH003        ----------------..---- ----.-------...--V...----..-----...---T-...------- -V-------T-----KN---...--.---....----.---
E_TH3.W2TH005        ----------------..---- ----.-------...-P-...----..-----...-----...------- --------K-----------...--K.---....----.---
E_TH3.W2TH006        ----------------..---- ----.-------...---...----..-----...-----...---Q--- ---E-----T-----K----...--.---....---X.---
E_TH3.W2TH007        -----V----------..---- ----.-------...-N-...----..-----...-----...------- --------K-----------...--.---....-M--.---
E_TH3.W2TH009        ----------------..---- ----.-------...---...----..-----...-----...------- ---R-----A-----K--E-...--.---....----.---
E_TH3.W2TH011        -------I--------..---- ----.-------...---...----..-----...-----...N------ -------Y-------K----...--T.---....----.---
E_TH3.W2TH015        ----------------..---- ----.------G...---...----..-----...-----...------- --------K-----X-----...--.---....----
E_TH3.W2TH018        ----------------..---- ----.-------...---...----..I----...-----...------- ---------A----------...--K.---....----
E_TH3.W2TH019        ----------------..-K-- ----.-------...---...----..-----...-----...-----H- -----E---A-----K----...--.---....----.---
E_TH3.W2TH020        ----------------..---- ----.-------...L--...----..-----...-----...------- ---------A----------...--.---....----.---
E_TH3.W2TH021        ----------------..---- ----.-------...L--...----..-----...-----...------- --------KA----------...--.---....----.---
E_TH3.W2TH022        ----------------..---- ----.-------...---...----..-----...-----...---R--- --------KA----------...--.---....----.---
E_TH3.W2TH023        ----------------..---- ----.-------...-N-...----..-----...-----...------- ------------K-------...--.---....----.---
E_TH3.W2TH024        ----------------..---- ----.-------...---...----..-L-K-...-----...------- --------K-----------...----R....----.---
E_TH4.D3TH966  ----------------------..---- ----.-------...T--...----..-----...---T-...N------ ---------A-----K----...--.---....----.------------....-----------------
E_TH4.D3TH975  -V---------Q----------..---- ----.-------...VR-...----..-----...-----...------- -------D-T---A------...--.---....----.-------K--....-----------------
E_TH5.CMU01         -----------------..---- ----.-------...---...----..-----...-E-T-...------- -----R---T---AK---G-...--.---....----.------------....------
E_TH5.CMU02         -----------------..---- ----.F.Y--I-...M-R...---H..-----...-K-V-...-P----- ---------A----------...--.---....----.------------....------
E_TH5.CMU03         -----------------..---- ----.----I--...--M...----..-----...---K-...------- --------------A-----...--T.---....-S--.------------....T-----
E_TH5.CMU04         -----------------..---- ----.A------...-P-...----..-L---...-----...N--S--- ---------T---AK-----...--.---....----.------P---....------
E_TH5.CMU05         -----------------..---- ----.-------...-R-...----..-----...-----...-P---F- -----E---A-----K----...--.---....-T-K.--------V-....------
```

V3 Region Alignments

```
                                                                _______________________ V3 LOOP ______________________
                       ^^^          ^^^     ^ *^    ^^^                                                  *   ^^^                       ^^^                              *      *
E_CONSENSUS_95 IIIRSENLTNNAKTIIVHLNKS..VEIN CTRP.SNNTRTS...ITI...GPGQ..VFYRT...GDIIG...DIRKAYC EINGTKWNEVLKQVTEKLKE...HFN.NKT....IIFQ.PPSGGDLEIT....MHHFNCRGEFFYCN
E_TH5.CMU06         --------I--------..---- ----.F--V--R...M-M...---H..-----...-E-T-...------- ----------T-----------...---.---....---.-----------....------
E_TH5.CMU07         ----D------------..---- ----.FRSI--R...TH-...-G--..AY---...KK---...---Q--- ----------T----A------...---.---....---.-----------....------
E_TH5.CMU08         -----------------..---- ----.--KK-QR...V--...---R..-Y-S-...---M-...-----H- ---------K------------...---.---....V--.-----------....------
E_TH5.CMU10         --------------HE-..---- ----.FY-K--R...TS-...-Q-R..-L---...---T-...N-G-P-- ---------K--N---------...----.-RN....S--.-----------....----I-
E_TH5.E1058         ---------------R-..---- ----.FKII---...TR-...---R..--HK-...-S-L-...------- ------N--K--R---K-----...--K.---....-T--.-----------....------
E_TH5.KH001         -----------------..---- ----.---I---...---...---R..R----...EG-G-...------- ----------------------...--E.--N....-S--.-----------....------
E_TH5.KH003         -----------------..---- ----.-------...-P-...-----..A----...-------...---R--- ---------K------G-----...---.---....----.-----------....------
E_TH5.KH004         ---------------R-..---- ----.F--IK-K...M-R...---H..---K-...-E-T-...------- ----------TI----------...--K.---....---R.-----------....------
E_TH5.KH005         -----------------..---- ----.TDI--I-...TR-...---R..---K-...-N-L-...------- ----------T----AG-----...---.--A....---K.-----------....------T
E_TH5.KH007         -----------------..---- ----.---R--K...--M...---R..-----...KE-V-...---Q--- ----------T-R--A---R-...--K.---....----.-----------....-----
E_TH5.KH008         -----------------..-A-- ----.--KI---...MR-...---R..--H--...-A-L-...------- ---------------K------...--K.---....----.-----------....------
E_TH5.KH009         ----D---I--------..---- ----.-K-I---...LR-...-----.-----...EG-N-...------- ---------K------------...---.---....V--.-----------....------
E_TH5.KH011         -----------------..---- ----..Y--KIR...--R...---R..-V---...-EMV-...------- ---------R----------V...---.---....N--.---
E_TH5.KH012         -----------------..---- ----.-------...--M...-----.-----...-----...------- -QV---I--KA-----------...--K.---....N--.---
E_TH5.KH013         -----------------..---- ----.FRKIS-R...TP-...-L-R..AI-K-...EN---...-----F- -----I---K--T---------...----.-SA....R--.---
E_TH5.KH014         -----------------..---- ----.F--K--R...M-M...---H..---K-...-E-T-...------- -K-------K------------...---.--K....----.---
E_TH5.KH015         -----------------..---- ----.FE-R--K...M-M...---H..-V-K-...---T-...------- ---------K---L-A----V...--I.---....----.---
E_TH6.0182     ----------------------..---- ----.-------...---...---R..-----...---T-...------- ----------------------...--K.---....---K.-----------....----------------
E_TH6.1012     ---------D-T----------..---- ----.---K--R...---...---R..-----...-E---...------- ---------K----------G...---.---....V--.-----------....----------------
E_TH6.1941     -------I--------------..---- ----.-------...-R-...-----.-----...-----...N------ ---------K------------...---.---....----.-----------....----S---------
E_UK1.11643    ----------------------..---- ----.---I---...FR-...-----.--HK-...-S-T-...------- ---------KA----A------...---.---....S--.-----------....----------------
```

V3 LOOP
F_CONSENSUS_95 IVVRSQNISDNTKTIIVHLNET..VQIN CTRP.NNNTRKS...IHL...GPGQ..AFYAT...GDIIG...DIRKAHC NVSGTQWNNTVQRVRAKLQS...HFP.NKT....ITFN.SSSGGDLEIT....MHSFNCRGEFFYCNT
F_BR.7944
F_BR.RJ103
F_BR1.BZ126
F_BR1.BZ162
F_BR1.BZ163
F_CM.CA16
F_CM.CA20
F_CM.CA4
F_GA.VI354
F_RO1.14018
F_RO1.14020
F_RO1.14024
F_RO1.14027
F_RO1.14028
F_RO1.14034
F_RO1.14036
F_RO1.14041
F_RO1.14046
F_RO2.RM53002
F_RO2.RM53011
F_RO2.RM53012
F_RO2.RM53013
F_RO2.RM53014
F_RO2.RM53015
F_RO2.RM53018
F_RO2.RM53021
F_RO2.RM53022
F_RO2.RM53023
F_RO2.RM53024
F_RO2.RM53027
F_RO2.RM53029
F_RO2.RM5303
F_RO2.RM53031
F_RO2.RM53032
F_RO2.RM53034
F_RO2.RM53035
F_RO2.RM53037
F_RO2.RM53040
F_RO2.RM53043
F_RO2.RM5306
F_RO2.RM5307
F_RO2.RM5308

V3 Region Alignments

A Subtype

At this time there are viral sequences from 139 HIV-1 infected individuals associated with HIV-1 subtype A. The A subtype consensus sequence (A_CONSENSUS_95) generated from these sequences was based on the most common amino acid found in each position in an alignment of these sequences. 120 of these sequences have been published and/or have been made available for printing in the database by their authors.

1) **CF1.ID#:** These twelve sequences are from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. Consensus, PCR-clones, cell culture, DNA. Murphy E, et al., *AIDS Res. Hum. Retroviruses* **9**:997–1006 (1993). GenBank accession numbers L11457–L11458, L11461–L11463, L11469–L11471, L11474–L11475, L11477–L11479, L11484–L11496, L11498, L11518, and L11523–L11524,
2) **CF2.GAN and SAS:** These sequences were kindly provided prior to publication by Dr. MP Kieny of Transgene, Strasbourg, France. They are a part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system. Year of isolation and health status of individuals from which the virus was isolated were not provided. Viruses were isolated in the CAR by C Mathiot and B You (Pasteur Inst., Bangui), grown by F Barre-Sinoussi and A Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D Schmitt and MP Kieny.
3) **CI2.CI-ID#:** These sequences are from 11 of 13 isolates from individuals from Abidjan, Cote d'Ivoire. CI-14 and CI-20 were symptomatic, and the others were asymptomatic. CI-14, CI-45 and CI-47 were serologically dually reactive for HIV-1 and HIV-2. The C2V3 region is part of a 900 bp sequence. Samples were collected between May 1990 and Sept. 1991. Virus was cultured with donor PBMCs, nested PCR amplified, 3–4 clones were sequenced, and the consensus of those clones is presented here. Janssens W, et al. *AIDS* **8**:21–26 (1994). GenBank accession numbers X72024–X72027, X72030–X72039, X72043–X72056, and X72059–X72065.
4) **CM1.CA-ID#:** These sequences are 11 of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic (CA7, CA11, CA15, CA17, CA18, and CA21) and symptomatic (CA1, CA2, CA6, CA19 and CA22) individuals. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate. Nkengasong JN, et al., *AIDS* **8**:1405–12 (1994). GenBank accession numbers for the entire set of 17 envelope sequences: X80438–X80454.
5) **DJ.DJ-ID#:** These three sequences from Djibouti were from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession numbers L22939, L22941, and L23064.
6) **GA.VI191:** This sequence from Gabon was from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22952.
7) **GH.D687:** A single sequence from an individual from Ghana, provided by Georg-Speyer Haus, Frankfurt, Germany (Dr. Ursula Dietrich). GenBank accession number L07652.
8) **KE.K89:** This sequence is named "KENYA" in the GenBank entry, but is identified as K89 in the original manuscript. It is a Kenyan sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22943.
9) **KE.KEN-ID#:** These 19 sequences were derived from patients who were part of a 1990–1992 cohort study of maternal risk factors in mother to child transmission, including 22 pregnant women and an infant from Kenya. The C2V3 region was sequenced. Janssens W et al., *AIDS Res. Hum.*

Retroviruses **10**:1577–1579 (1994). GenBank accession numbers for the entire set of 23 patients surveyed in this study: U12984–U13006.

10) **NG.NI:** A full gp120 envelope sequence from Nigeria was kindly provided by Dr. Tom Howard from the University of Southern California (USC) Howard et al. *AIDS Res. Hum. Retroviruses* **10**:1755–1757 (1994). No GenBank entry has yet been created for this sequence.
11) **RW.564C** This sequence represents 10 identical sequences generated from PCR amplified plasma RNA from one of three infants in a Dutch mother/infant study. Patient pair 564 was from Rwanda. A sample was collected from the infant at 30 months of age. Samples were also collected from the mother at 12 and 30 months after the birth. Mother sequences are not included in this consensus. [Mulder-Kampinga et al.(1993)]. [Mulder-Kampinga et al.(1995)]. The child 564 Env sequence is from the entry with GenBank accession number Z47881. Mother 564 sequenes are in entries with GenBank accession numbers Z47882–Z47902. Mother sequences are not included in this alignment. Gag gene sequences from mother/child pairs are also available in Genbank accession numbers Z47903–Z47911; Z47912–Z47928; Z47929–Z47935; Z47936–Z47950. The second mother/child pair was from the Netherlands, see G_NL.127C. The third mother/infant pair in this study was from the Netherlands, see B_NL.114C.
12) **RW.SF1703:** This sequence is from Rwandan isolate sf170, a biologically active clone reported to be macrophage-tropic. Cheng-Mayer C, Homsy J, Evans LA, and Levy J. *Proc. Natl. Acad. Sci. USA*, **85**:2815–19 (1988); and, Evans, L, Higgins, D, Cheng-Mayer C and Levy J. *Virology* **181**:288–294 (1991). GenBank accession number M66533.
13) **RW.W2RW-ID#:** Eight sequences from asymptomatic individuals from Rwanda sampled in 1992. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are: [Wolf et al.(1994)]; [Osmanov et al.(1994)]; [Gao et al.(1994a)]. GenBank accession numbers U08630–U08641, U08647–U08665, U08763–U08766, and U08793–U08794.
14) **RW2-ID#:** Nine consensus sequences from Rwanda. Saah, A. Unpublished 1994. GenBank accession numbers U23216–U23373.
15) **TZ.TAN-ID#:** These four sequences are from a set of 14 Tanzanian samples from symptomatic individuals, using serum samples taken in 1988 to generate PCR clones from viral RNA for sequencing. Zwart G, et al., *AIDS* **7**:467–474 (1993). GenBank accession numbers L01313, L01315–L01316, L01335, L01337–L01339.
16) **TZ2.ID#:** These two sequences were from patients at a clinic in Dar es Salaam, Tanzania. The individuals from which the virus was cultured showed clinical signs of AIDS, and the year of viral isolation was 1988. Viral cDNA was PCR amplified from donor PBMC, and one cloned PCR product per donor was sequenced. [Siwka et al.(1994)]. GenBank accession numbers U12408, U12409.
17) **UG.1033:** This sequence is a consensus sequence of blood and CSF samples taken from a Ugandan patient 1033, CDC class IV-A. sequences which compose this consensus are: Z23182–Z23184, Z23220 Keys et al., *Virology* **196**:475–483 (1993). GenBank accession numbers Z23177, Z23182–Z23184, and Z23220–Z23223.
18) **UG.964:** A single sequence used in a study of the impact of sequence variation on the distrinution and seroreactivity of linear antigenic epitopes. The sequence was derived from PCR amplified DNA from peripheral blood leukocytes. The patient was an asymptomatic individual from Uganda. [Pestano et al.(1995)]. GenBank accession number U11599. cf. B_US17.ID#, C_UG1.45, and D_UG7.ID#.
19) **UG.U455:** This sequence is from the 1985 Ugandan isolate U455; the complete genomic sequence is available. Oram JD, et al., *AIDS Res. Hum. Retroviruses* **6**:1073–1078 (1990). GenBank accession number M62320.
20) **UG.UG06:** This sequence is from a Ugandan isolate. Atkin A, et al., *AIDS Res. Hum. Retroviruses* **9**:351–356 (1993). GenBank accession number M98503.
21) **UG1.W2UG-ID#:** Three sequences from asymptomatic individuals from Uganda in 1992. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region

WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are: [Wolf et al.(1994)]; [Osmanov et al.(1994)]; [Gao et al.(1994a)]. GenBank accession numbers U08666–U08669 and U08767–U08770, U08788–U08792, U08795, U09124 and U09127

22) **UG2-ID#:** These 11 sequences are part of a set of sequences derived from 22 Ugandans who were attending an AIDS clinic, sampled in 1990. Consensus, PCR-clones, peripheral blood DNA. Albert J, et al., *Virology* **190**:674–681 (1992). GenBank accession numbers M98902–M98905, M98908–M98910, M98914–M98917, M98919, M98924–M98928, M98938–M98941, M98946–M98966, and M98976–M98978,
23) **UG4.UG-ID#:** Two Ugandan sequences from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession numbers L22957 and L22951.
24) **ZR.Z321:** This sequence is from the 1976 Zairean isolate Z321. Srinivasan A, et al., *AIDS Res. Hum. Retroviruses* **5**:121–9 (1989). GenBank accession number M15896.
25) **ZR1.ID#:** These ten sequences are part of a set of 14 A and D sequences from women from Zaire; 8 were healthy, 4 showed minor signs of illness, and 2 had AIDS. PCR-direct, peripheral blood DNA. Potts KE, et al., *AIDS Res. Hum. Retroviruses* **9**:613–618 (1993). GenBank accession numbers L19624–L19626, L19628–L19630, L19632–L19634, and L19636.
26) **??.BLR10A:** This sequence is from GenBank accession number L38411, Lukashov,V.V. Unpublished.
27) **??.5393:** This sequence is from GenBank accession number L07082, Bex,F. et al. Unpublished.

The A subtype sequences which are not yet published, and for which the authors have not yet given permission for release, are from the following set:

1) **CI1.ID#:** Ten sequences from individuals from the Cote d'Ivoire. PCR-direct, peripheral blood DNA. These sequences were provided by the Centers for Disease Control, Atlanta GA, USA (Dr. Chin-Yih Ou and Dr. Marcia Kalish).
2) **UG3.ID#:** Six sequences from individuals from Uganda. PCR-clone, peripheral blood DNA. These sequences were provided by the Centers for Disease Control, Atlanta GA, USA (Dr. Chin-Yih Ou and Dr. Marcia Kalish).
3) **ZR2.ID#:** Three sequences from individuals from Zaire, provided by the LTCB, NCI, NIH, Bethesda, MD, USA (Dr. Marvin Reitz).

B Subtype

At this time we have included viral sequences from 519 HIV-1 infected individuals associated with HIV-1 subtype B. The B subtype consensus sequence (B_CONSENSUS_95) generated from these sequences was based on the most common amino acid found in each position in an alignment of these sequences. Please note that none of the studies which have published sequences of only the V3 loop sequences are included here, as the DNA sequences were deemed too short for phylogenetic analyses. (For example, LaRosa G, et al., *Science* **249**:932–935 (1990) and Fouchier RAM, et al., *J. Virol.* **66**:3183–3187 (1992).) 503 of the 519 sequences have been published and/or have been made available for printing in the database by their authors.

1) **AU1.ID#:** Put forth as evidence that coinfection by multiple HIV-1 strains can occur in vivo, these three concensus (MRC1,MRC2, and MRC3) come from an Austrailian homosexual male who had been infected by more than one sexual partner, and harbored three distict strains of HIV-1 B. The authors also found recombinant sequences, not included here. The sequences were PCR amplified from plasma RNA and PBMC DNA. [Zhu et al.(1995)]. GenBank accession numbers U16372–U16388.
2) **BE.SIMI84** One of two cloned env sequences from a patient with AIDS from Belgium. A vaccinia construct that expresses this gene was created to vaccinate the patient's non-infected brother with the goal of immune therapy by adoptive transfer of lymphocytes. [Bex et al.(1994)]. GenBank accession number L07421.
3) **BR1.ID#:** These 21 sequences represent the B env subtype sequences found among 22 Brazilian outpatients with varying degrees of disease progression. Consensus, PCR clones, peripheral blood PBMC DNA. Potts KE, et al., *AIDS* **7**:1191–1197 (1993). GenBank accession numbers for 20 of the 56 clones from which consensus sequences were calculated: L19225–L19236, L19240–L19246 and L20963.
4) **BR2.W2BR-ID#:** 13 sequences from individuals from Brazil. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are in press: de Wolf F, et al. *AIDS Res. Hum. Retroviruses* **10**:1387–1400 (1994); Osmanov S, et al. *AIDS Res. Hum. Retroviruses* **10**:1325–26 (1994); [Gao et al.(1994a)]. GenBank accession numbers U08670–U08714, U08771–U08778, U08780–U08782, U08792, U08796, U08797–U08798 and U08800,
5) **BR3.RJ- or SP-ID#:** These 19 sequences came from the Brazilian cities Rio de Janeiro and Sao Paulo. The sequences that are very short, containing V3 loop fragments insufficient for phylogenetic analysis, are not included here (5 of the 26). They included 19 viral sequences are of the B subtype. An F subtype and a B-F recombinant were also observed in this set. Year of isolation for the sequences range from 1990–1992 for Rio de Janeiro, and 1992 for Sao Paulo. The only two with CD4+ cells < 200 were RJ636 and RJ27. The CDC clinical class ranged from II-IV. DNA extracted from PBMCs of HIV infected individuals was amplified, and the PCR product was directly sequenced. Morgado et al., *AIDS Res. Hum. Retroviruses* **10**:569–576 (1994) and Sabino EC et al. *J. Virol.* **68**:6340–6346 (1994). GenBank accession numbers U00400–U00401, U00403, U00405, U00407–U00414, U00416–U00418, U00421, U00424–U00425, and U00427.
6) **BR4.BZ-ID#:** These 2 sequences are from seropositive Brazilian patients. Virus was cultured on donor PBMCs and proviral DNA was harvested from possitive cultures. PCR was used to generate sequencing templates. [Louwagie et al.(1994)]. GenBank accession numbers L22087 and L22088. The gag gene sequences from these same isolates are also available in L11752 and L11754. cf F_BR2.BZ-ID#.
7) **BR5.ID#:** These 10 sequences are from entries with GenBank accession numbers L19328–L19337. Bandea, CI. Unpublished.
8) **CH.ID#:** These 10 sequences came from 24 individuals living in Geneva, Switzerland who were recently infected at the time of blood drawing. Samples were collected between January 1988 and September 1993. Sequences were determined directly from PCR products of uncultured PBMC DNA or serum cDNA. All subjects were asymptomatic, 19 subjects had p24 antigen levels

ranging from 5 to 6,357 pg/ml and 5 subjects had no detectable p24 antigen. Two subjects were epidemiologically linked (K11 and K16) so only one of those two is presented here. Two other individuals showed identical DNA sequences over the entire V3 region (K53 and K77) so only one of them is presented here. Three other individuals (K13, K42 and P4) had sequences nearly identical to the LAI or IIIB laboratory strains of HIV. Although the authors are convinced that these are not contaminants, and that a IIIB-like strain of HIV is circulating in Geneva, they are not included in this alignment. [Antonioli et al.(1995)]. GenBank accession numbers U10957–U10980.

9) **CI.CI-22:** A single B subtype sequence from a set of 13 isolates from individuals from Abidjan, Cote d'Ivoire. CI-22 was symptomatic. The C2V3 region is part of a 900 bp fragment that was sequenced for each individual. Samples were collected between May 1990 and Sept. 1991. Virus was cultured with donor PBMCs, nested PCR amplified, 3 clones were sequenced, and the consensus of those clones is presented here. Janssens W, et al. *AIDS* **8**:21–26 (1994). GenBank accession numbers X72040–X72042.
10) **CM.CA5:** A single B subtype sequence from a set of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic and symptomatic HIV infected individuals, specifically, patient CA-5 was asymptomatic. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate. Nkengasong JN, et al., *AIDS* **8**:1405–12 (1994). GenBank accession numbers for the full set of 17 sequences which includes CA5 are X80438–X80454.
11) **DE.D31:** This sequence is from isolate D31. Kreutz A, et al., *AIDS Res. Hum. Retroviruses* **8**:1619–1629 (1992). GenBank accession number X61240.
12) **DE.HAN:** This sequence is from an infectious clone from the German isolate DE.HAN-2. Sauermann U, et al., *AIDS Res. Hum. Retroviruses* **6**:813–823 (1990). GenBank accession number U43141.
13) **FR.J61:** This sequence is from one of the JBB clones from the French patient Bru. Wain-Hobson S, et al. *Science* **252**:961–64 (1991) and Guo et al. *Nature* **349**:745–6 (1991). GenBank accession numbers M64178–M64223, M64406–M64415 and X57449–X57459
14) **FR.LAI:** This sequence is from the French isolate LAI (formerly BRU) which is also referred to as IIIB. Wain-Hobson S, Sonigo P, Danos O, Cole S, and Alizon M. *Cell* **40**:9–17 (1985). Also see: Alizon M, Wain-Hobson S, Montagnier L, and Sonigo P, *Cell* **46**:63–74 (1986) and Wain-Hobson S, et al. *Science* **252**:961–964 (1991). GenBank accession numbers K02013, L23090–L23103, X01762, L48389. Other sequences which are of this type include: PV22, K02083; MFA, M33943; un-named, Z11530; BH8, K02011; BH10, M15654; TH4, L31963; and HXB, K03455 M38432 M14100. This isolate of HIV-1 has also been extensively studied in cases such as the infected lab worker. See for example [Reitz et al.(1994)] U12030–U12055.
15) **GA.OYI:** This sequence is from the Gabonese isolate OYI (designated elsewhere as isolate 397), isolated from a healthy HIV-1 infected individual. GA.OYI appears to have been the first viral sequence from Africa that phylogenetically clustered with North American viruses. Huet T, et al., *AIDS* **3**:707–715 (1992). GenBank accession number M26727.
16) **HT.RF:** This sequence is from the clone HAT-3, from Haitian isolate RF. Starcich BR, et al., *Cell* **45**:637–648 (1986). GenBank accession number M17451.
17) **HT1.D-ID#:** These seven sequences are from Haitians, and are part of a set of sequences generated as part of the DAIDS variation program in the laboratories of Dr. Beatrice Hahn at the University of Alabama, and Dr. Marcia Kalish at the the Centers for Disease Control, Atlanta, GA. Except for D2HA590, the full gp160 was sequenced from clones derived from expanded culture stocks. D2HA590 is a direct sequence from PCR amplified DNA from expanded culture. The sequence ID numbers are abbreviated, for example D2HA590 can be read as DAIDS sequence (D), isolated in 1992 (2), Haitian (HA), patient 301590 (590). GenBank accession numbers: U08441–U08447, U04900. Both U08441 and U08442 are sequences from patient HT1.D1HA651 and are identical over the region of interest. GenBank accession numbers for additional clones derived from these patients: U04901–U04906.
18) **HT2.H-ID#:** These 25 sequences are from Haitians. All sequences were PCR amplified from the infected individuals PBMCs, and this set includes direct sequences of PCR amplification products, consensus sequences of multiple clones of PCR products plus one direct sequence, and single clones

of PCR products. These sequences were provided by the Centers for Disease Control, Atlanta, GA USA (Dr. Chin-Yih Ou), and John Hopkins University School of Hygiene and Public Health, Balimore, MD USA (Dr. Neal Halsy), and the Centers for Development and Health, Complexe Medico Sociale de la Cite Soleil, Port-au-Prince, Haiti (Dr. Reginald Boulos). GenBank accession numbers L07145–L07161, L07163–L07165, L07167–L07207, L07209–L07239, L07241–L07246.

19) **IN.IN9:** This sequence was isolated in India. Tripathy,S.P., Renjifo,B., Wang,W., McLane,M., Ostermann,J., Bollinger,R., Rodrigues,J.J., Tripathy,S.P. and Essex,M. unpublished. National Aids Research Institute, 73 'G' Block, MIDC, Bhosari, Pune 411 026, India GenBank accession number U31364. cf C_IN3.ID#.
20) **IN1.ID#:** These four sequences were isolated in Hyderabad, Andhra Pradesh, in southern India. The C2V3 region of env was amplified by nested priming from DNA from PBLs from fresh blood samples. Date of sampling and health status of HIV-1 infected individuals is unknown. Baskar PV, et al., *AIDS Res. Hum. Retroviruses* **10**:1039–1041 (1994). GenBank accession numbers L29091–L29094.
21) **IT.ID#:** These two sequences are consensus sequences from 4 clones each. They were obtained from PCR-amplified proviral DNA from Langerhans cells from skin patches from a deceased AIDS victim in Italy [Sala et al.(1995)]. Small V1-V2 region sequences and V3-loop sequences from the same skin samples were published in [Sala et al.(1994)]. GenBank accession numbers U20670–U20677 used here. GenBank Accession numbers Z34376–Z34458, Z34470–Z34513 and Z34515 were V1-V2 and V3-loop sequences from the same patient.
22) **IT1.ID#:** These 10 sequences are from infants infected in utero. The sequences came from PCR amplified DNA of uncultured PBMCs, PCR amplified DNA of cultured PBMCs, or from RNA from serum collected at or shortly after delivery. [Scarlatti et al.(1993)]. GenBank accession numbers L08277–L08286. Sequences from the mothers of theses infants are also available in entries with accession numbers L08287–L08372.
23) **JP.GUNA:** A Japanese 1989 isolate HIVGUN, infectious to T cells, was adapted to growth on fibroblast-like BT cells. A single amino acid change at the tip of the V3 loop was shown to be responsible for the change in tropism, GPGR to GSGR. Takeuchi Y, Akutsu M, Murayama K, Shimizu N, and Hoshino H. J. Virol. **65**:1710–1718 (1991). GenBank accession number M59192.
24) **JP.JH32:** This is a sequence from a lambda clone of Japanese isolate JH3. Komiyama N, et al., *AIDS Res. Hum. Retroviruses* **5**:411–419 (1989). GenBank accession number M21138.
25) **NL.114C** This consensus sequence represents sequences generated from PCR amplified plasma RNA from one of three infants in a Dutch mother/infant study. Samples were collected from the infant at birth, at 6 weeks and at 9 months of age. Samples were also collected from the mother before birth, at birth and after birth. Mother sequences are not included in this consensus. [Mulder-Kampinga et al.(1993)]. [Mulder-Kampinga et al.(1995)]. Infant 114 is from GenBank accession numbers L21111–L21153. Mother 114 sequences are from GenBank accession numbers L21028–L21110. Infant 127 sequences are from GenBank accession numbers Z47817–Z47832. Mother 127 sequences are from GenBank accession numbers Z47833–Z47880. Gag gene sequences from mother/child pairs are also available in Genbank accession numbers Z47903–Z47911; Z47912–Z47928; Z47929–Z47935; Z47936–Z47950. The second mother/child pair was also from the Netherlands, see G_NL.127C. The third mother/infant pair in this study was from Rwanda, see A_RW.564C.
26) **NL.168:** This is a consensus sequence of 3 clones after culturing in PBMC. The isolate was originally from an AIDS patient in Amsterdam. [Wrin et al.(1989)]. GenBank accession numbers U15030–U15032. A V3-loop (105 bp) segment from the original isoalte has been previously reported. [Fouchier et al.(1992)]. GenBank accession number L06694.
27) **NL.X1:** This is a consensus sequence of 10 clones from a recipient in a donor-recipient study. Sequences from donor Y and recipient X2 are also part of this study, but are not included here. [Cornelissen et al.(1995)]. GenBank accession numbers Z47505–Z47514 are from X1. Other new sequences analyzed in this paper include Z47411–Z47540. Sequences M91828–M91838 (donor H and recipient O referred to as patients A14 and A13, respectively in [Wolfs et al.(1992)] see B_NL1.A13) were also re-analyzed in this study.

28) **NL1.ID#:** These nine sequences are part of a study of presumed donor-recipient pairs from an HIV-1 transmission study conducted in the Netherlands. If pairs were extremely close or identical, only the "recipient" is included here. Recipient samples were from the first sample to be antibody positive, and are numbers 1,3,5,7,9, and 13. These sequences are consensus sequences of multiple clones from PCR amplified serum RNA. Wolfs TFW, Zwart G, Bakker M, and Goudsmit *J. Virology* **189**:103–110 (1992). GenBank accession numbers M91819–M91827, M91829, M91831–M91832, M91839, M91857–M91870, M91872, M91874, M91881–M91884, M91891, M91893, 91895–M91908, M91910, M91911–M91926. Number 13, and the donor were also analyzed in [Cornelissen et al.(1995)].
29) **NL2.ID#:** These two sequences are part of a Dutch study of mutations occurring over a five year period (starting in 1985) in two patients. Serum RNA was PCR amplified and multiple clones were sequenced. The consensus for each patient is shown. Wolfs TFW, Zwart G, Bakker M, Valk M, Kuiken CL, and Goudmsit *J. Virology* **185**:195–205 (1991). GenBank accession numbers M74591–M74684.
30) **NL3.NET-ID#:** These six consensus sequences from the Netherlands are samples from AIDS patients, using serum samples to generate PCR clones from viral RNA for sequencing. Zwart G, et al., *AIDS* **7**:467–474 (1993). GenBank accession numbers L01282–L01297.
31) **NL4.ID#:** These 74 sequences represent a study of early seroconverters from different times with different risk factors for transmission during the AIDS epidemic in the Netherlands. The year the sample was taken is indicated in the last part of the sequence name. The risk group of the individual from whom the virus is derived is indicated in the first letter of the sequence name (I, B and H for IVDUs, hemophiliacs, and homosexuals, respectively). Viral genomic RNA from sera was PCR amplified and amplification product was direct sequenced. Kuiken CL, et al., *Proc. Natl. Acad. Sci. USA* **90**:9061–9065 (1993). GenBank accession numbers Z29219–Z29225, Z29256–Z29258, and Z29262–Z29325.
32) **NL5.ID#:** These 18 consensus sequences are from a study of patients with, and without, AIDS dementia complex (ADC) in the Netherlands. Not all patients were from the Netherlands. Samples were collected between 1986 and 1992. Viral genomic RNA from sera and/or cerebral spinal fluid was reverse transcribed and PCR amplified and clones were sequenced. [Kuiken et al.(1995)] GenBank accession numbers Z37734–Z37963, Z37970–Z37971
33) **NL6.ID#:** 16 is a consensus sequences of four sequences used in a study of HIV-1 envelope-mediated syncytium formation. The consensus represents four clones from one patient, two clones of the consensus are SI and two are NSI. 320 is a single SI clone. The sequences were derived from PCR amplified DNA from provirus cultured in MN cells. [Andeweg et al.(1992)]. GenBank accession numbers L08655–L08662.
34) **NL7.ID#:** These two consensus sequences are from sets of sequences (GenBank accession numbers U13240, U13241, U13243–U13247 for consensus 537 and U13242, U13248–U13252 for 1058) used in a study on the dynamics of HIV sequence changes in vivo and the utility of heteroduplex analysis. Both sequences were derived from PCR amplified PBMC DNA. Concensus 537 represents a set of sequences from a Dutch patient with a relatively stable CD4+ cell count at 62 months post-seroconversion. Concensus 1058 represents sequences from another Dutch patient whose CD4+ cell count at 73 months post-seroconversion was declining faster than 537's. [Delwart et al.(1994)]. cf. US10.ID#.
35) **NL8.ID#:** These two consensus and two individual sequences came from patients early in infection, before, or around the time of seroconversion. #537 amd #1058 are consensus sequences. [Shpaer et al.(1994)]. GenBank accession numbers U23651–U23663, U23667, U23670. cf. US11.ID#.
36) **PR1.D-ID#:** These four sequences are from Puerto Rico, and were generated as part of the DAIDS variation program in the laboratory of Dr. Marcia Kalish at the the Centers for Disease Control, Atlanta, GA. The C2V3 region was directly sequenced from PCR amplification of DNA from viral culture. The sequence ID numbers are abbreviated; for example D2PR732 can be read as DAIDS sequence (D), isolated in 1992 (2), Puerto Rico (PR), patient 301732 (732). A full description of these sequences can be found in the April 1994 supplement to the HIV database, part III. GenBank accession numbers U04926–U04929.

37) **PY.ID#:** Ten sequences from 10 patients living in Asuncion, Paraguay. All 10 were male patients with symptoms of AIDS. Virus was propagated in tissue culture for an unstated length of time prior to harvesting proviral DNA for PCR and sequencing. PCR products were directly sequenced. PY.3614, PY.3615 and PY.12837 were syncytium-inducing and the other 7 were not. The tenth sequence, PY.3614p was PCR amplified directly from patient PBMCs with no culturing. The significant differences between the cultured sequence from this patient (PY.3614c) and the direct sequence, indicates that the virus that grew out in culture was a minority of the virus present in PBMC. The sequence of PY.3614c is not included in this alignment. [Cabello et al.(1995)]. GenBank accession numbers U28949–U28959.
38) **SE1.ID#:** Seven sequences that are consensus sequences of blood and CSF samples taken from each patient. The CDC disease stage class for the patients are as follows: II pts 930, 2815; III pt 931; IV-E pt 2951; IV-A pt 1032; and IV-C2 pts 1433, 1866. Keys et al., *Virology* **196**:475–483 (1993). GenBank accession numbers Z23178–Z23181, Z23185–Z23187, Z23192–Z23194, Z23200–Z23219, Z23224–Z23227, Z23232–Z23235, and Z23240–Z23255.
39) **TH.T8174:** This sequence comes from a study of the genetic heterogeneity and epidemiological distribution of HIV1 in Thailand. The host was an intravainous drug user and the sequence was obtained from PCR amplified PBMC DNA. [Ou et al.(1993)]. GenBank accession number L19238. cf. E_TH.T8178.
40) **TH1.ID#:** These ten sequences are from individuals from Thailand. PCR-direct, peripheral blood PBMC DNA. Referred to as Thai subtype B in Ou et al. Ou C-Y et al., *AIDS Res. Hum. Retroviruses* **8**:1471–1472 (1992) and Ou C-Y et al. Lancet **341**:1171–1174 (1993). (Published erratum appears in Lancet **342**:250 (1993).) GenBank accession numbers L07442, L07449–L07456 and L07460.
41) **TH2.ID#:**: The TB132 sequence is from a set of isolates from HIV seropositive individuals from Thailand. PCR, PBMC co-culture, DNA. Full env sequence is available. McCutchan FE, et al., *AIDS Res. Hum. Retroviruses* **8**:1887–1895 (1992). Please note: the TB132 locus name in the database corresponds to the McCutchan et al.'s "BK132" isolate. GenBank accession number L03697. The CM237 sequence is from DNA from PBMC. [Mascola et al.(1994)]. GenBank accession number L14570. See also B_US14
42) **TH3.W2TH-ID#:** 2 sequences from Thailand from asymptomatic individuals. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the *Human Retroviruses and AIDS 1993* database. Relevant papers are in press: de Wolf F, et al. *AIDS Res. Hum. Retroviruses* **10**:1387–1400 (1994); Osmanov S, et al.*AIDS Res. Hum. Retroviruses* **10**:1325–26 (1994); [Gao et al.(1994a)]. GenBank accession numbers U08715–U08718, U08801–U08802, and U08783–U08784.
43) **TH4.ID#:** These twelve sequences are B subtype sequences from Thailand. Ten were genetically most similar to HIV-1 found in the Americas and Europe; these sequences were derived from people infected prior to 1988 (diagnosed in 1986 or 1987). The other two (N762 and N763) were designated B' and were isolated from people with more recent infections, 1988 and 1992. The sequences were obtained from PCR amplified PBMC DNA. The naming of the sequences submitted to GenBank does not correspond with the naming of the sequences in the paper. Kalish ML, et al, *AIDS Res. Hum. Retroviruses* **10**:1573–1575 (1994). GenBank accession numbers for the entire set of thirteen sequences studied in this publication: U15576–U15588.
44) **TH5.ID#:** These three sequences are B subtype sequences from Thailand. Two individuals believed to be dually infected with subtypes B and E were analyzed. It is not clear from the paper or the GenBank entries, which sequences came from individual 1 and which from 2. [Artenstein et al.(1995)]. Genbank accession numbers U21471, U21473, U21475. See also E_TH6.ID#.
45) **UK.CAM1:** This sequence is from the British isolate CAM1. McIntosh A, and Karpas A, Thesis (1991), Cambridge University, England. GenBank accession number D10112.
46) **UK.V-ID#:** A set of six sequences from a study of hemophiliacs from Scotland who were originally thought to have been infected by the same batch of factor VIII. (ScV12 is a sequence from a hemophiliac from the U.S., included as a control). All are consensus sequences of multiple direct PCR sequences obtained from limiting dilution of PBMCs. The Scottish hemophiliacs were infected in 1984 and the PBMCs were obtained for analysis in 1989. Although the samples were potentially

related, they were deemed sufficiently divergent in this region for inclusion in this set. Simmonds P, Balfe P, Ludlam CA, Bishop JO, and Leigh Brown, AJ. *J. Virol.* **64**:5840–5850 (1990), and Balfe P, Simmonds P, Ludlam CA, Bishop JO, and Leigh Brown, AJ. *J. Virol.* **64**:6221–6233 (1990). GenBank accession numbers M61327–M61346 and M61391–M61407. GenBank entries with accession numbers M84240–M84317 are more sequences from patient 82, taken over the period from 1984–1991 [Holmes et al.(1992)]. GenBank entries with accession numbers L13488–L13497 are also from these patients [Zhang et al.(1993)]

47) **UK1.CPHL1:** This is a consensus from the British isolate 93–08020, clones 1, 4, 7, 18, 19 and 43. It was referred to as 93–08020 in [Arnold et al.(1995c)] and was isolated from the patient referred to as CPHL1 in [Arnold et al.(1995a)]. CPHL1 is a surgeon and CPHL2 was a patient of his in 1986, approximately 7 years prior to sampling for this study. Because sequences from CPHL1 and CPHL2 are no more similar to each other than to sequences from the general population, transmission cannot be concluded, and both sequences are included in this alignment. GenBank accession numbers U21100 (clone 1) and U23112–U23116 (clones 18, 19, 4, 43 and 7 respectively).
48) **UK2.CPHL2:** This is a consensus from the British isolate 93–17305, clones 3, 11, 18 and 25. It was isolated from the patient referred to as CPHL2 in [Arnold et al.(1995a)]. GenBank accession numbers U23117–U23120 (clones 11, 18, 25 and 3 respectively).
49) **UK3.CPHL7:** This sequence is a sequence from the British isolate 94–24612, clone 13. It was isolated from the patient referred to as CPHL7 in [Arnold et al.(1995a)]. GenBank accession number U23126. U23127 is a second clone from this same isolate. Sequences from three other patients epidemiologically linked to CPHL7 (CPHL6, accession numbers U23130–U23132; CPHL8, U23128–U23129; CPHL9, U23133–U23135) are not included in this alignment.
50) **US.ACP1:** This virus was cultured from a seronegative man with Kaposi's sarcoma. (See: Ho D, The American J. of Med., **86**:349–351 (1989). ACP1 was the sequenced after one passage in PBMCs. The sequence AC-H9 (GenBank M80661) was also derived from this patient. Ashkenazi, A, et al., *Proc. Natl. Acad. Sci., USA* **88**:7056–7060 (1991). GenBank accession number M80660.
51) **US.ADA:** This sequence is from the monocytropic U.S. isolate ADA. Westervelt P, Gendelman HE, and Ratner L, *Proc. Natl. Acad. Sci. USA* **88**:3097–3101 (1991). GenBank accession number M60472.
52) **US.ALA1:** This sequence is from an infectious clone of the 1985 U.S. isolate AL-1, taken from a patient with AIDS. Buckler-White A, Theodore T, et al., Unpublished (1988). GenBank accession number M38430.
53) **US.BAL1:** This sequence is from the macrophage tropic U.S. isolate BAL, harvested from lung alveolar tissue. Reitz M, et al., Unpublished (1990). GenBank accession number M68893.
54) **US.BCSG3:** This is a fragment of a full genomic sequence from the provirus SG3, cloned as a single proviral unit. This clone replicates more efficiently in chimpanzeee than in human lymphocytes, and is extremely cytopathic in immortalized human T-cell lines. Ghosh SK, et al., Virology 194: 858–864 (1993). GenBank accession number L02317.
55) **US.BRVA:** This sequence is from U.S. isolate BRVA, which was taken from the brain tissue of a AIDS patient with neurological disorders. Anand R, et al., *Virology* **168**:79–89 (1989). GenBank accession number M21098.
56) **US.BWB:** This consensus sequence is from sequences derived from PCR amplified PBMC DNA from brain tissue. [Monken et al.(1995)]. GenBank accession numbers L17088–L17126.
57) **US.CDC42:** This sequence is from an infectious clone of the U.S. isolate CDC-451. Desai SM, et al., *Proc. Natl. Acad. Sci. USA* **83**:8380–8384 (1986). GenBank accession number M13137.
58) **US.JFL:** This sequence is from a non-infectious clone from the monocytropic U.S. isolate JFL. McNearney T, et al., *Proc. Natl. Acad. Sci. USA* **87**:1917–21 (1990). GenBank accession number M31451. Other sequences from HIV-1 isolates epidemiologically linked to this isolate can be found in GenBank entries with accession numbers L06256–L06273. [McNearney et al.(1993)].
59) **US.JM:** This sequence, along with B_US.WM, came from viral isolates after short term culture in PBMCs, PCR amplification, and cloning of PCR products. Both are from asymptomatic, seropositive individuals. Ashkenazi, A, et al., *Proc. Natl. Acad. Sci., USA* **88**:7056–7060 (1990). GenBank accession number M80662.

60) **US.JRCSF:** This sequence is from an infectious clone of the 1986 U.S. isolate JRCSF, derived from from the CSF of a patient who died with Kaposi's sarcoma and severe AIDS encephalopathy. The infectious clone JRFL was isolated from the brain of the same patient. O'Brien WA, et al., *Nature* 348, 69–73 (1990). Also see: Pang S, Vinters HV, Akashi T, O'Brien WA, and Chen ISY, *J AIDS* **4**:1082–1092 (1991). GenBank accession number M38429.
61) **US.MA:** This is a consensus of 35 sequences obtained either after viral culture or from PCR amplification products from patient PBMCs. Patient MA, from the US, was infected in 1984 or 1985, and had been experiencing neurological disorders prior to 1989 when the sample was taken. Kusumi K, et al., *J. Virol.* **66**:875–885 (1992). GenBank accession numbers M79342–M79354 and M90025–M90046.
62) **US.MN:** This sequence is from 1984 U.S. isolate MN, taken from a pediatric AIDS patient. Gurgo C, et al., *Virology* **164**:531–536 (1988). GenBank accession number M17449. See also GenBank entries with accession numbers U15030–U15032.
63) **US.NY5CG:** This sequence is from the 1984 U.S. isolate NY5. Willey R.W., et al., *Proc. Natl. Acad. Sci. USA* **83**:5038–5042 (1986). GenBank accession number M38431. See also GenBank accession number K03346.
64) **US.P896:** This sequence represents a molecular clone from an primary isolate derived from a Jamaican man who immigrated to Philidelphia 15 years earlier. At the time of viral isolation, he had no antiviral therapy, but was an AIDS patient with $<$ 10 CD4 cells per mm3. The infectious molecular clone from which this sequence was derived is both macrophage-tropic and extremely cytopathic in lymphocytes. Collman R, et al. *J. Virol.* **66**:7517–21 (1992). GenBank accession number M96155.
65) **US.RJS:** This is a consensus of six biologically characterized clones from isolate RJS, isolate 4. The HIV-1 infected individual had been infected for five years at the time of isolation. Complete env sequence is available. Daniels RS, Smith MH, and Fisher AG. *J. Virol.* **65**:5574–5578 (1991) and Fisher AG et al., *Nature* **334**:444–447 (1988). GenBank accession numbers M37491 and M37573–M37577.
66) **US.SB(A-C):** These three sequences are from 1988 U.S. isolates taken from a woman, her daughter and her sexual partner. The three viruses are epidemiologically linked, however the amino acids sequences appeared sufficiently divergent in this region to merit the inclusion of all three samples. Sequences were directly sequenced from PCR amplification products after the virus was briefly cultured. Burger H, et al., *Proc. Natl. Acad. Sci. USA* **88**:11236–11240 (1991). GenBank accession numbers M77228–M77230.
67) **US.SC:** This sequence is from the 1984 U.S. isolate SC, from an AIDS patient. Gurgo C, et al., *Virology* **164**:531–536 (1988). GenBank accession number M17450.
68) **US.SF162:** This sequence is from an infectious clone from the U.S. isolate SF162, cultured from cerebrospinal fluid. Cheng-Mayer C, Quiroga M, Tung JW, Dina D, and Levy J, *J. Virol.* **64**:4390–4398 (1990). GenBank accession number M38428.
69) **US.SF2:** This sequence is from an infectious clone from the U.S. isolate ARV-2. Sanchez-Pescador R, et al., *Science* **227**:484–492 (1985). HIVSF13 (GenBank accession number L07422) is a more infectious virus taken from the same patient [Cheng-Mayer et al.(1991)]. GenBank accession number K02007.
70) **US.SF33:** This sequence is from an infectious clone from the 1984 U.S. isolate SF33. York-Higgins D, Cheng-Mayer C, Bauer D, Levy JA, and Dina D, *J. Virol.* **64**:4016–4020 (1990). GenBank accession number M38427.
71) **US.TN-ID#:** These eight sequences are from asymptomatic individuals identified after donating blood in Memphis, Tennessee, USA. [Slobod et al.(1994)]. GenBank accession numbers U09140–U09171.
72) **US.WM:** This sequence, along with B_US.JM, came from viral isolates after short term culture in PBMCs, PCR amplification, and cloning of PCR products. Both are from asymptomatic, seropositive individuals. Ashkenazi, A, et al., *Proc. Natl. Acad. Sci., USA* **88**:7056–7060 (1990). GenBank accession number M80663.

73) **US.WMJ22:** This sequence is from the isolate WMJ22, isolated from a person of Haitian descent living in the U.S. Hahn BH, et al., *Science* **232**:1548–1553 (1986). GenBank accession number M12507.
74) **US.YU:** This is a consensus sequence of eight lambda phage clones and 12 PCR amplified clones derived from the uncultured brain tissue of a patient with AIDS dementia complex. A macrophage tropic clone (YU-2) is almost identical to the consensus sequence of YU in this region, with only a single amino acid change (K to N) 10 amino acids from the carboxy-terminal end of the sequence. Li Y, et al., *J. Virol.* **65**:3973–3985 (1991). Complete genomic sequences are available for two of the HIV1YU clones, along with biological characterizations of four of the HIV1YU clones: Li Y, et al., *J. Virol.* **66**:6587–6600 (1991). The GenBank accession numbers for the YU-10 and YU-2 complete genomes are M93259 and M93258, respectively.
75) **US1.HC-ID#:** These are forty 1990–1991 U.S. samples, from the study of the dentist who was thought to have been the source of HIV-1 infection of six of his patients. Only the dentist's viral sequence and the Florida control sequences are shown here; the six epidemiologically and genetically linked patients are excluded from this alignment as their viral sequences were very similar to the dentist's. All sequences were PCR amplified from patient PBMCs. Most are direct sequences from the amplification products, although some are consensus sequences of multiple clones of PCR products and one direct sequence, Ou, C-Y, et al., *Science* **256**:1165–1171 (1992). GenBank accession numbers for 70 of 75 sequences in this set: M90847–M90853, M90881–M90886, M90894–M90900, M90914–M90956, and M90958–M90964.
76) **US2.D-ID#:** These 15 sequences are from the USA, and are part of a set of sequences generated as part of the DAIDS variation program in the laboratory of Dr. Marcia Kalish at the the Centers for Disease Control, Atlanta, GA. The C2V3 region was directly sequenced from PCR amplification products of DNA from viral culture. The sequence ID numbers are abbreviated, for example D2US711 can be read as DAIDS sequence (D), isolated in 1992 (2), United States (US), patient 301711 (711). A full description of these sequences can be found in the April 1994 supplement to the HIV database, part III. GenBank accession numbers U04907–U04915, U04918 and U04921–U04925.
77) **US3.D-ID#:** These four sequences are from the US, and are part of a set of sequences generated as part of the DAIDS variation program in the laboratory of Dr. Beatrice Hahn at the University of Alabama. gp160 sequences of clones from expanded culture stocks are available. The sequence ID numbers are abbreviated, for example D2US711 can be read as DAIDS sequence (D), isolated in 1992 (2), United States (US), patient 301711 (711). A full description of these sequences can be found in the April 1994 supplement to the HIV database, part III. GenBank accession numbers U08448–U08449 and U08451–U08452. GenBank accession numbers for additional clones from these patients: U04916–U04917 and U04919–U04920,
78) **US4.ZhuPt-ID#:** Sequences from five primary seroconverters. Consensus, PCR-clones, peripheral blood PBMC DNA. Although sequences were also available from two of the donors, only sequences from five recipients are shown here. Zhu et al., *Science* **261**:1179–1181 (1993). GenBank accession numbers L21224–L21324, L21372–L21373, L21376–L21377, L21380–L21381, L21384–L21389, L21393–L21394, L21397–L21398, L21400–L21401, L21404, L21405, L21408, L21410–L21411, L21414, L21417–L21418, L21426–L21427, L21430–L21432, L21434–L21437, L21440–L21442, L21445, L21447, L21449–L21450, L21453–L21454, L21457, L21459, L21462–L21463, L21519–L21520, L21522–L21523, L21525–L21526, L21528–L21533, L21535–L21536, L21538–L21539, L21541, L21543, L21545–L21546, L21548, L21550–L21551, L21553–L21554, L21556, L21558–L21559, L21561–L21569, L21571–L21578, L21580, L21582, L21584, L21586, L21588, and L21590.
78) **US5.pt-ID#:** A consensus of PCR-clones, peripheral blood DNA. These two sequences were part of a study of blood sequences compared to brain sequences from six individuals. Korber BTM, et al., *J. Virol.* **68**:7467–7481 (1994). GenBank accession numbers U05360–U05568.
79) **US6.-ID#:** Each sequence is a consensus sequence of several cloned PCR products from PBMC proviral DNA from an individual infant. These sequences were part of a study of mother-infant transmission. Infant blood samples were taken from 1 week (infant 7) to 34 months (infant 4) post-partum. Dates of sample collection were: Infant 1, 10/25/91; Infants 2–4, 10/31/91; Infant

5, 2/6/92; Infant 6, 8/30/93; Infant 7 5/13/93. Maternal sequences were also reported as part of this set, but are not used in building these consensus sequences. [Ahmad et al.(1995)]. GenBank accession numbers U16390–U16652.

80) **US7.-ID#:** Each sequence is a consensus sequence of several cloned PCR products from PBMC proviral DNA from an individual patient. These sequences were part of a study of early samples (1984–1986) from the San Francisco region. One of the samples (552–3) was HIV negative but was determined to be contaminated with blood from another (565–3) so the six samples from 552–3 and 565–3 are made into one consensus here (565). [Sabino et al.(1994)]. GenBank accession numbers L20371–L20381.

81) **US8.-ID#:** These sequences were from a study of three recipients of contaminated blood. Recipient 1 (R1) and recipient 2 (R2) each recieved blood from different donors. A third recipient, not presented here, recieved blood from both donors. All three recipients were neonates. R1 recieved erythrocytes from donor 1 on 19 October, 1984 at the age of 3.5 weeks. For R1 the sequence of one of the two clones (2E) is presented here. R2 recieved erythrocytes from donor 2 on 24 September, 1984 at the age of 2 months. For R2 a consensus of 6 clones is presented here. Blood samples for this study were drawn in March 1986. R1 had slow weight gain, and R2 had lymphadenopathy at time of sample collection. [Diaz et al.(1995)]. GenBank accession numbers U11188 = R1–2E, U11189 not used; U11203, U11196, U11199, U11192–U11194 = R2 six clones. The tat and envelope V4-V5 regions of clones from these same individuals are also available in U11173–U11178, U11180, U11205–U11209.

82) **US9.-ID#:** These are consensus sequences for samples taken over a range of time from four diferent subjects. Blood samples for S1 were drawn in Nov '85, Jul '87, Jan '88 and May '89. Blood samples for S2 were drawn in May '85, Apr '87 and Oct '87. Blood samples for S3 were drawn in Jun '87 and Dec '87. Blood samples for S4 were drawn in Jan '85, Jan '89 and Jun '89. S2, S3 and S4 had decreasing CD4 counts during the study period. S1 had fluctuating CD4 counts. [McNearney et al.(1992)]. GenBank accession numbers L03430–L03453 and L23575–L23588 = S1; L03454–L03477 and L23618–L23633 = S2; L03478–L03490 and L23589–L23600 = S3; L03491–L03515 and L23601–L23617 = S4.

83) **US10.ID#:** These three consensus sequences are from sets of sequences used in a study on the dynamics of HIV sequence changes in vivo and the utility of heteroduplex analysis. All sequences were derived from PCR amplified PBMC DNA. The MA145 consensus represents sequences (GenBank accession numbers U00821, U00822, U00831–U00839) taken from an asyptomatic male from Massachusetts over a period of 4.5 years starting April 1989. The SFBU and SFPE consensuses represent sequences (GenBank accession numbers U13373–U13380 and U13381–U13388 respectively) taken from two patients with AIDS from San Fransisco. [Delwart et al.(1994)]. cf. NL7.ID#.

84) **US11.ID#:** These four consensus and seven individual sequences came from patients early in infection, before, or around the time of seroconversion. Sequences for 306, 419, 349, and 074 are consensus sequences. [Shpaer et al.(1994)]. GenBank accession numbers U23664–U23666, U23668, U23669, U23671–U23708. cf. NL9.ID#.

85) **US12.ID#:** These sequenses were used in an investigation into the transmission of HIV-1 from one child (CHA), who had recieved zidovudine, to another child (CHB), who harbored a zidovudine-resistant strain. The presence of the zidovudine-resistant strain in child 1 and 2, and the lack of such a strain in child 2's mother was used to show that child 2 was infected by child 1 and not by child 2's mother. LC sequences are from children used as local controls. All sequences were derived from PCR amplified PBMC DNA. [Fitzgibbon et al.(1993)]. GenBank accession numbers L12752–L12756, L19695, L19697. L12756 is listed as "isolate 100" in GenBank, but seems to be the "group B consensus sequence" used for phylogenetic analysis.

86) **US13.ID#:** These three consensus sequences are from three IV drug users in Florida. Proviral DNA sequences were obtained from blood, cerebrospinal fluid and dorsal root ganglia from each of the three individuals. Sequences for V1-V5 of env, were PCR amplified, cloned and sequenced. [Shapshak et al.(1995)]. [Xin et al.(1995)]. The sequence for patient 149 is a consensus of all 24 clones in GenBank accession numbers U16094–U16117. The sequence for 141 is a consensus of GenBank entries labelled as being from patient 141 with the exception of: R5D, R6D, R7D, R2D

and R4D, which were similar to IIIB strains of HIV-1; and R1R, R3R, R7R, R8R and R9R, which were similar to samples from patient 144. The sequence for 144 is a consensus of GenBank entries labelled as being from patient 144 with the exception of: R3R, R6R, R9R, R12R, R13R and L1D, which were similar to IIIB strains of HIV-1; and C3D, C4D, C7D, C8D and C10D, which were similar to samples from patient 141. While infection of each individual with multiple strains of HIV including one very similar to the IIIB lab strain, is a possible explanation of these findings, we are only including one consensus sequence from each patient for this alignment. GenBank accession numbers for patients 141 and 144 are U16032–U16093.

87) **US14.ID#:** These four sequences are from DNA from PBMC. [Mascola et al.(1994)]. GenBank accession numbers L14573–L14576. See also B_TH2, E_TH2.
88) **US15.ID#:** These six consensus sequences are from a study of infants. Blood samples were collected from six infants over time. Strunnikova,N., Ray,S.C., Livingston,R., Rubalcaba,E. and Viscidi,R.P. unpublished. Hopkins University School of Medicine, Baltimore, MD. GenBank accession numbers U22682–U22810, U22834, U22835.
89) **US16.ID#:** These two consensus sequences were used in a study of the impact of sequence variation on the distrinution and seroreactivity of linear antigenic epitopes. Both sets of sequences were from PCR amplified DNA from peripheral blood leukocytes. Patient ARTC1 was an asymptomatic individual from New York and ARTC3 was an AIDS patient from New York. [Pestano et al.(1995)]. GenBank accession numbers U11586–U11594. cf. A_UG1.964, C_UG1.45, and D_UG7.ID#.
90) **US17.ID#:** CB7 is a consensus sequence for 12 clones from two different samples (six clones each) from the same patient, collected in 1988 and 1990, plus two more clones as yet unpublished. The patient seroconverted in 1985. The patient did not receive any antiviral therapy until 1992. The patients CD4 count was 1035 in 1988 and 807 in 1990. The patient's PBMS were cocultured with donor PBMC for an unspecified length of time before cultured DNA was isolated and PCR amplified. Individual clones of PCR product were them sequenced. [Wang et al.(1995)]. GenBank accession numbers U16324–U16335 and U19706–U19711. The other 8 sequences are individual clones from 8 different patients, all from the same clinic in Boston, MA. Individual CB7 was again included in this study, submitted for publication to *J. Virol.* (1995). GenBank accession numbers U19694–U19711.
91) **VE.ID#:** These 8 sequences are from 8 individuals in Venezuala. Quinones-Mateu,M.E., Dopazo,J., Rota,T.R., Acevedo,N., Este,J.A. and Domingo,E. unpublished, Univeridad Autonoma de Madrid, Spain. GenBank accession numbers U16764–U16778.
92) **??.ID#:** These 7 sequences are from unpublished GenBank entries by Lukashov,V.V. with accession numbers: L38405, RUS3A; L38407, RUS4A; L38417, LIT18A; L38412, LIT11A; L38419, LIT21A; L38416, LIT17A; L38420, LIT17A.

The 16 B subtype sequences which are not yet published, and for which the authors have not yet given permission for release, are from the following sets:

1) **PE:** A single envelope sequence from Peru, provided by the Centers for Disease Control, Atlanta, GA, USA, and the Walter Reed Research Institute.
2) **US.FO and FR.J61** A sequence from the United States, and one from France, provided by Dr. Marvin Reitz of the NIH.
3) **US:** Six consensus consensus of PCR-clones, from peripheral blood DNA. These sequences were provided by Dr. Steven Wolinsky of Northwestern University Medical School, Chicago, IL, USA.
4) Two sequences of unspecified origin provided by Dr. Karen Coates Fryer, Macfarlane Burnet Centre for Medical Research, Fairfield, Victoria, Australia.
5) Five sequences provided by Dr. Beatrice Hahn (BZ01, IZ01, WE101, USHOBR and USFASH).

C Subtype

At this time there are viral sequences from 57 HIV-1 infected individuals associated with HIV-1 subtype C. The C subtype consensus sequence (C_CONSENSUS_95) generated from these sequences was based on the most common amino acid found in each position of an alignment. 55 of these sequences have been published and/or have been made available for printing in the database by their authors.

1) **BR.W2BR025:** This sequence is part of a gp160 sequence from an asymptomatic individual from Brazil, sampled in 1992. A clone was derived from an expanded viral culture, expressed and sequenced. This sequence was provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences from this patient can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are: De Wolf F, et al. *AIDS Res. Hum. Retroviruses* **10**:1387–1400 (1994); Osmanov S, et al. *AIDS Res. Hum. Retroviruses* **10**:1325–1326 (1994); [Gao et al.(1994a)]. GenBank accession number U15121. Entries with accession numbers U09126, U09132 and U09133 are also from W2BR025.
2) **CF1.15166:** A single sequence from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. Consensus, PCR clones, cultured virus, DNA. Murphy E, et al., *AIDS Res. Hum. Retroviruses* **9**:997–1006 (1993). GenBank accession number L11525. Other sequences from this study were subtypes A and E.
3) **DJ1.DJ-ID#:** These two sequences from Djibouti were from a a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession numbers L22940 and L23065.
4) **IN.IN#:** These 8 sequences sequence were isolated in India. Tripathy,S.P., Renjifo,B., Wang,W., McLane,M., Ostermann,J., Bollinger,R., Rodrigues,J.J., Tripathy,S.P. and Essex,M. unpublished. National Aids Research Institute, 73 'G' Block, MIDC, Bhosari, Pune 411 026, India GenBank accession numbers U29179, U29694–U29698, U31362 and U31363. cf B_IN.IN9.
5) **IN1.D-ID#:** These four sequences are from samples from high risk patients in India, PCR clones, DNA, PBMC culture. Dietrich U et al., *AIDS* **7**:23–27 (1993). GenBank accession numbers L07651 and L07653–L07655; X65638–X65640 and X68406.
6) **IN2.D-ID#:** These five sequences are from samples from high risk patients in India, primarily stage I. They were nested PCR amplified from DNA obtained from uncultured PBMC from patients serologically defined as HIV-1/HIV-2 mixed infections. Grez M et al., J Virol 68: 2161–8 (1994). GenBank accession numbers U07098 and U07100–U07103.
7) **IN3.ID#:** These eight sequences were from Pune and New Delhi, India. Unpublished (1995) Tripathy,S.P., Renjifo,B., Wang,W., McLane,M., Ostermann,J., Bollinger,R., Rodrigues,J.J., Tripathy,S.P. and Essex,M. National Aids Research Institute, 73 'G' Block, MIDC, Bhosari, Pune 411 026, India GenBank accession numbers U29179, U29694–U29698, U31362, U31363. cf B_IN.ID#.
8) **MW1.ID#:** These 15 sequences are from pregnant women with risk factors from Malawi. PCR-direct, peripheral blood DNA. Orloff GM et al., *AIDS Res. Hum. Retroviruses* **9**:705–706 (1993). GenBank accesssion numbers L07427–L07441, L15721–L15733 and L15735.
9) **MW2.D-ID#:** These two sequences are from individuals from Malawi, generated as part of the DAIDS variation program in the laboratory of Dr. Beatrice Hahn at the University of Alabama. The C2V3 was excised from full gp160 sequences, derived from clones from expanded culture stocks. The sequence ID numbers are abbreviated, for example D3MA959 can be read as DAIDS sequence (D), isolated in 1993 (3), Malawi (MA), patient 301959 (959). GenBank accession numbers U08453–U08454.
10) **SN.SE364:** A Senegalese sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22944.

11) **SO.1574:** This sequence is a consensus sequence of blood and CSF samples taken from the Somalian patient 1574, CDC classification II. Keys et al., *Virology* **196**:475–483 (1993). GenBank accession numbers Z23188, Z23190–Z23191, and Z23228–Z23231.
12) **SO.SM145:** A Somalian sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22946.
13) **UG.45:** A single sequence used in a study of the impact of sequence variation on the distrinution and seroreactivity of linear antigenic epitopes. The sequence was derived from PCR amplified DNA from peripheral blood leukocytes. The patient was an asymptomatic individual from Uganda. [Pestano et al.(1995)]. GenBank accession number U11597. cf. A_UG1.964, B_US17.ID#, and D_UG7.ID#.
14) **UG.UG268:** A Ugandan sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22948.
15) **UK1.00513:** This sequence is from the British isolate 93–00513. [Arnold et al.(1995c)]. GenBank accession number U21099.
16) **ZA.NOF:** This sequence is from a South African individual, was part of a study of HIV-1 strains in India, this sequence was found to be closer than other isolates from Africa. PCR amplified DNA from PBMC cultures were sequenced. Dietrich U et al., *AIDS* **7**:23–27 (1993). GenBank accession numbers L07426 and U06716. cf. C_IN.ID#:
17) **ZM1.ZAM-ID#:** Two sequences from Zambia, from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession numbers L22954 and L22956.
18) **ZW.2647:** This sequence is a consensus sequence taken from blood and CSF samples taken from Zimbabwe patient 2647, CDC classification II. Keys et al., Virology **196**:475–483 (1993). GenBank accession numbers Z23196–Z23199 and Z23236–Z23239.
19) **??.ID#:** These 7 sequences are from unpublished GenBank entries by Lukashov,V.V. with accession numbers: L38418, RUS20A; L38404, RUS2A; L38406, RUS1A; L38410, BLR9A; L3841, RUS13A; L38409, BLR8A; L38408, BLR5A.

The two C subtype sequences which are not yet published, and for which the authors have not yet given permission for release, are from S. Africa, and Zambia, provided by the Dept. of Med. Virology, University of Stellenbosch, Tygerberg, South Africa (Dr. Michael Becker).

D Subtype

At this time there are viral sequences from 89 HIV-1 infected individuals associated with HIV-1 subtype D. The D subtype consensus sequence (D_CONSENSUS_95) generated from these sequences was based on the most common amino acid found in each position of an alignment. All but 7 of these sequences have been published and/or have been made available for printing in the database by their authors.

1) **CF.4020:** This sample was the only D subtype from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. Consensus, PCR-clones, cell culture, DNA. Murphy E, et al., *AIDS Res. Hum. Retroviruses* **9**:997–1006 (1993). A full gp120 sequence from this isolate was kindly provided prior to publication by Dr. MP Kieny of Transgene, Strasbourg, France. It is a part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system. Year of isolation and health status of individuals from which the virus was isolated were not provided. Viruses were isolated in the CAR by C Mathiot and B You (Pasteur Inst., Bangui), grown by F Barre-Sinoussi and A Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D Schmitt and MP Kieny. GenBank accession numbers L11472–L11473.
2) **CI.CI-13:** A single D subtype sequence from a set of 13 isolates from individuals from Abidjan, Cote d'Ivoire. CI-13 was symptomatic, and serologically dually reactive for HIV-1 and HIV-2. The C2V3 region is part of a 900 bp fragment that was sequenced for each individual. Samples were collected between May 1990 and Sept. 1991. Virus was cultured with donor PBMCs, nested PCR amplified, 3–4 clones were sequenced, and the consensus of those clones is presented here. Janssens W, et al. *AIDS* **8**:21–26 (1994). GenBank accession numbers X72028–X72029.
3) **KE1.KEN-ID#:** These three patients were part of a 1990–1992 cohort study of maternal risk factors in mother to child transmission, including 22 pregnant women and an infant from Kenya. The C2V3 region was sequenced. W Janssens et al., in press *AIDS Res. Hum. Retroviruses* (1994). GenBank accession numbers for sequences from the entire set of 23 patients studied in this publication: U12984–U13006.
4) **SN.SE365:** A Senegalese sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22945.
5) **TZ.TAN-ID#:** These ten sequences are from from a set of 14 Tanzanian samples from symptomatic individuals, using serum samples taken in 1988 to generate PCR clones from viral RNA for sequencing. Zwart G, et al., *AIDS* **7**:467–474 (1993). GenBank accession numbers L01298–L01339.
6) **TZ2.ID#:** These eight sequences were from patients at a clinic in Dar es Salaam, Tanzania. The individuals from which the virus was cultured showed clinical signs of AIDS, and the year of viral isolation was 1988. Viral cDNA was PCR amplified from donor PBMC, and one cloned PCR product per donor was sequenced. [Siwka et al.(1994)]. GenBank accession numbers U12406, U12407, U12410–U12415.
7) **UG.U44342:** This sequence is from a Ugandan. Consensus of PCR-clones, peripheral blood DNA. Intact env sequences are available from this sample. Bruce C, C Clegg, A Featherstone, J Smith and J Oram. *AIDS Res. Hum. Retroviruses* **9**:357–368 (1992). GenBank accession numbers M98408–M98416.
8) **UG.UG23:** This sequence is from a Ugandan isolate. Atkin A, et al. and Pestano G, et al., unpublished. Andrew Atkin is at CUNY, NY, NY, USA. GenBank accession number M98504.
9) **UG1.W2UG-ID#:** Twelve sequences from asymptomatic individuals from Uganda in 1992. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are in press: de Wolf F, et al. *AIDS Res. Hum. Retroviruses* (1994); Osmanov S, et al. *AIDS Res. Hum. Retroviruses* (1994); Gao F, et al. *AIDS Res. Hum. Retroviruses*

(1994). GenBank accession numbers HU08721–U08741, HU08786–U08787, U08803–U08809 and U08821–U08824.

10) **UG2.ID#:** These twelve sequences are from 1986–1987 Ugandan samples. Consensus, PCR-clones, peripheral blood DNA. Oram JD, et al., *AIDS Res. Hum. Retroviruses* **7**:605–614 (1991). GenBank accession number for one of the 161 sequences used: M62320.
11) **UG3.ID#:** These 11 sequences are part of a set of sequences derived from 22 Ugandans who were attending an AIDS clinic, obtained in 1990. Consensus, PCR-clones, peripheral blood DNA. Albert J, et al., *Virology* **190**:674–681 (1992). GenBank accession numbers L00614–L00618, L00733–L00737, M98894–M98899, M98901, M98906–M98907, M98911–M98913, M98918, M98920–M98923, M98929–M98937, M98942–M98945, and M98967–M98975.
12) **UG5.UG-ID#:** Three Ugandan sequences from a set of HIV-1 viral isolates from Africa. Health status of the individuals from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession numbers L22947 and L22949–L22950.
13) **UG6.ID#:** Three Ugandan sequences from a set of HIV-1 viral isolates from Africa. All three individuals from which the virus was cultured had AIDS, and the year of viral isolation was 1987. Viruses were cultured with HUT-78 cells for an unspecified length of time. The V3 region of env (gp160) was amplified, cloned and sequenced. [von Brunn et al.(1995)]. GenBank accession numbers U15005, U15006 and U15007.
14) **UG7.ID#:** These sequences were used in a study of the impact of sequence variation on the distrinution and seroreactivity of linear antigenic epitopes. Both sets of sequences were from PCR amplified DNA from peripheral blood leukocytes. All patients were asyptomatic individuals from reporting for regular blood drawing at the Nakasero blood transfusion service, Kampala, Uganda. [Pestano et al.(1995)]. GenBank accession numbers U11595, U11596, and U11598. cf. A_UG1.964, C_UG1.45, and D_UG7.ID#.
15) **UK1.CPHL4:** This sequence is a consensus from the British isolate 93–43424, clones 2, 6 and 35. It was referred to as 93–43424 in [Arnold et al.(1995c)] and as CPHL4 in [Arnold et al.(1995a)]. GenBank accession numbers U21098 (clone 35) and U23121–U23122 (clones 2 and 6 respectively). CPHL4 is a female who is believed to have contracted the virus from CPHL5 through heterosexual contact. Sequences from CPHL5 are not included in this alignment due to this epidemiological relationship.
16) **US.AMK:** This sequence comes from a student living in Alabama, who moved from Zaire to the US in 1988. Virus was isolated from the patients PBMCs; this isolate was PCR amplified, and amplification products from both gag and env were subcloned and sequenced. His CD4 count was < 5 cells/mm3, and he was symptomatic at the time of viral isolation. [Gao et al.(1994b)]. GenBank accession number U08193.
17) **ZR.ELI:** This sequence is from the Zairean isolate ELI. Alizon M, Wain-Hobson S, Montagnier L, and Sonigo P. *Cell* **46**:63–74 (1986). The complete genomic sequence and an infectious clone are available. GenBank accession number M27949.
18) **ZR.JY1:** This sequence is from Zairean isolate Z-84, clone JY1. Yourno J, et al., *AIDS Res. Hum. Retroviruses* **4**:165–173 (1988). GenBank accession number J03653.
19) **ZR.MAL:** This sequence is from a non-infectious clone of the Zairean isolate MAL. Alizon M, Wain-Hobson S, Montagnier L, and Sonigo P. *Cell* **46**:63–74 (1986). The complete genomic sequence and an infectious clone from the isolate MAL are available. MAL is known to be recombinant between subtypes A and D. GenBank accession number K03456.
20) **ZR.NDK:** This sequence is from an infectious clone of the Zairean isolate NDK. The molecular clone is highly cytopathic in vitro. Spire B, et al., *Gene* **81**:275–284 (1989). The complete genomic sequence is available. GenBank accession number M27323.
21) **ZR.Z2Z6:** This sequence is from an infectious clone of Zairean isolate Z2. Theodore T, and Buckler-White A, unpublished, and Srinivasan A, et al., *Gene* **52**:71–82 (1987). The complete genomic sequence is available. GenBank accession number M22639. See also [Srinivasan et al.(1987)] GenBank entry with accession numbers K03458 and M16322, which is from the same isolate.

22) **ZR1.ID#:** These four sequences are part of a set of 14 A and D sequences from women from Zaire. 8 were healthy, 4 showed minor signs of illness, and 2 had AIDS. PCR-direct, peripheral blood DNA. Potts KE, et al., *AIDS Res. Hum. Retroviruses* **9**:613–618 (1993). GenBank accession numbers L19623, L19627, L19631 and L19635.
23) **??.ID#:** This sequence is from an unpublished GenBank entry by Lukashov,V.V. with accession number: L38415, RUS14A.

The seven D subtype sequences which are not yet published, and for which the authors have not yet given permission for release, are from the following set:

UG4.ID#: Seven sequences from individuals from Uganda. PCR-direct, peripheral blood DNA. These sequences were provided by the Centers for Disease Control, Atlanta GA, USA (Dr. Chin-Yih Ou).

E Subtype

At this time there are viral sequences from 77 HIV-1 infected individuals associated with HIV-1 subtype E. The E subtype consensus sequence (E_CONSENSUS_95) generated from these sequences was based on the most common amino acid found in each position of an alignment. All 77 of these sequences have been published and/or have been made available for printing in the database by their authors.

1) **CF1.ID#:** These eight sequences are from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. Consensus, PCR-clones, cell culture, DNA. Murphy E, et al., *AIDS Res. Hum. Retroviruses* **9**:997–1006 (1993). GenBank accession numbers L11459–L11460, L11463–L11468, L11476, L11480–L11481, L11504–L11505, L11511–L11513 and L11519–L11521.
2) **CF2.ID#:** These three sequences were kindly provided prior to publication by Dr. MP Kieny of Transgene, Strasbourg Cedex, France. It is a part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system. Year of isolation and health status of individuals from which the virus was isolated were not provided. Viruses were isolated in the CAR by C Mathiot and B You (Pasteur Inst., Bangui), grown by F Barre-Sinoussi and A Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D Schmitt and MP Kieny.
3) **CM.CA10:** A single E subtype sequence from a set of 17 sequences from a very diverse set of isolates from from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic and symptomatic HIV infected individuals, specifically, CA10 was symptomatic. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate. Nkengasong JN, et al., *AIDS* **8**:1405–12 (1994). GenBank accession numbers for the entire set of 17 sequences in this publication: X80438–X80454.
4) **TH.T8178:** This sequence comes from a study of the genetic heterogeneity and epidemiological distribution of HIV1 in Thailand. The host was a female prostitute and the sequence was obtained from PCR amplified PBMC DNA. [Ou et al.(1993)]. GenBank accession number L19239. cf. B_TH.T8174.
5) **TH.N764** This sequence is a E subtype sequence from Thailand (THP13). 12 of 13 sequences from Thai prisoners were of subtype B; N764 represents the only subtype E sequence identified in this set, from a prisoner infected in 1989. The sequences were obtained from PCR amplified PBMC DNA. Kalish ML, et al, *AIDS Res. Hum. Retroviruses* **10**:1573–1575 (1994). GenBank accession number U15588.
6) **TH1.ID#:** These twelve sequences are from a set of 23 individuals from Thailand. PCR-direct, peripheral blood PBMC DNA. Referred to as Thai subtype A in Ou et al. Ou C-Y et al., *AIDS Res. Hum. Retroviruses* **8**:1471–1472 (1992) and Ou C-Y et al. *Lancet* **341**:1171–1174 (1993). (Published erratum appears in *Lancet* **342**:250 (1993).) GenBank accession numbers L07443–L07445, L07447–L07448, L07457–L07459 and L07461–L07464.
7) **TH2.ID#:** Six of these eight sequences are from 16 isolates from HIV seropositive individuals from Thailand. PCR, PBMC co-culture, DNA. Full env sequence is available. McCutchan FE, et al., *AIDS Res. Hum. Retroviruses* **8**:1887–1895 (1992). Please note: the "TN-ID#" locus names in the database correspond to the McCutchan et al.'s "CM-ID#" isolates. GenBank accession numbers L03698–L03701 and L03703–L03704. The other two (TH238, TN240) are also from Thailand, DNA from PBMC. [Mascola et al.(1994)]. GenBank accession numbers L14571, L14572.
8) **TH3.W2TH-ID#:** Fifteen sequences from asymptomatic individuals from Thailand in 1992. Consensus, PCR-clones, cell-culture DNA and RNA. These sequences were provided by the WHO Global Programme on AIDS Vaccine Development Study. The complete set of C2V3 region WHO sequences can be found in the April supplement to the Human Retroviruses and AIDS 1993 database. Relevant papers are: De Wolf F, et al. *AIDS Res. Hum. Retroviruses* **10**:1387–1400 (1994); Osmanov S, et al. *AIDS Res. Hum. Retroviruses* **10**:1325–1326 (1994); [Gao et al.(1994a)]. GenBank accession numbers U08810–U08811, U08825–U08836 and U08742–U08761. Entry with accession number U09131 is also TH_W2TH022.

9) **TH4.D-ID#:** These two sequences are are part of a set of sequences generated for the DAIDS variation program in the laboratory of Dr. Beatrice Hahn at the University of Alabama. They are clones from expanded culture stocks, and are excised from full gp160 sequences. The sequence ID numbers are abbreviated, for example D3TH966 can be read as DAIDS sequence (D), isolated in 1993 (3), Thai (TH), patient 301966 (966). GenBank accession numbers U08456 and U08457. The entry with accession number U08458 is also Thai.
10) **TH5.ID#:** These seventeen consensus and five individual sequences are from twenty two patients with AIDS involved in a study of genotypic and phenotypic characteristics of Thai HIV-1. Blood samples were collected between July and December 1993. All sequences were derived from PCR amplified PBMC DNA, after patient PBMCs were cocultured with virus-free donor PBMCs. CMU01, CMU03, CMU04, CMU05, CMU07, and CMU10 are NSI, the rest are SI, as determined by syncytium formation in the cocultured cells. CM = Chaing Mai University Hospital. KH = Kavila Army Hospital. All subjects were males and reported past contact with commercial female sex workers, but no history of drug injection, blood transfusion or homosexual contact. [Yu et al.(1995)]. GenBank accession numbers U25550–U25626.
11) **TH6.ID#:** These three sequences are E subtype sequences from Thailand. Two individuals believed to be dually infected with subtypes B and E were analyzed. It is not clear from the paper or the GenBank entries, which sequences came from individual 1 and which from 2. [Artenstein et al.(1995)]. Genbank accession numbers U21472, U21474, U21476.. See also B_TH5.ID#.
12) **UK1.11643:** This sequence is from the British isolate 94–11643. The sequence was determined from PCR-amplified lymphocyte DNA. The gag gene from this isolate was subtype A. The patient is thought to have contracted the virus in Thailand, but currently lives in the United Kingdom. [Arnold et al.(1995c)]. GenBank accession number U21109.

NOTE:

1) While the sequences in this subtype were distinct over this region of env from the other four env subtypes, in the gag gene it is not possible to make a distinction between this subtype and subtype A. What this means is that the isolates for which both gag and env are sequenced which cluster together as the “A” subtype in gag, are very distinctive in env and are broken down into two subtypes. env “A” and env “E”. This holds true for the E subtypes sequences that originated in Thailand, as well as the E subtype isolate from the Central African Republic for which gag sequence was obtained. McCutchan FE, et al., *AIDS Res. Hum. Retroviruses* **8**:1887–1895 (1992); Louwagie J, et al., *AIDS* **7**:769–780 (1993); and Murphy E, et al., *AIDS Res. Hum. Retroviruses* **9**:997–1006 (1993).
2) The relative lack of diversity in the Thai sequences in this subtype relative to the other subtypes is likely to be a consequence of the short time span of the HIV-1 subtype E epidemic in Thailand. McCutchan FE, et al., *AIDS Res. Hum. Retroviruses* **8**:1887–1895 (1992), and Ou Y-C et al., *AIDS Res. Hum. Retroviruses* **8**:1471–1472 (1992).

F Subtype

At this time there are viral sequences from 42 HIV-1 infected individuals associated with HIV-1 subtype F. The F subtype consensus sequence (F_CONSENSUS_95) generated from these sequences was based on the most common amino acid found in each position of an alignment. All of these sequences have been published.

1) **BR.7944:** This sequence represents a single env F subtype sequence found among 22 Brazilian outpatients with varying degrees of disease progression. It was identified by Potts et al. as the single sequence which did not cluster with North American sequences in phylogenetic analysis. Consensus, PCR clones, peripheral blood PBMC DNA. Potts KE, et al., *AIDS* **7**:1191–1197 (1993). GenBank accession number L19237.
2) **BR.RJI03:** An F subtype sequence from Rio de Janeiro, Brazil. 26 additional B and a B-F recombinant were also observed in this set. Year of isolation for was 1993, from an individual of CDC clinical stage II. Morgado et al., *AIDS Res. Hum. Retroviruses* **10**:569–576 (1994). DNA was amplified directly PBMCs of HIV infected individual, and the PCR product was directly sequenced. GenBank accession number U00422.
3) **BR1.BZ-ID#:** Three sequences from Brazil of the F subtype. Full length env (gp160) was amplified from provial DNA of cultured PBMCs, cloned and sequenced. [Louwagie et al.(1994)]. GenBank accession numbers L22082, L22084 and L22085. The gag gene of these same isolates is found in L22083, L22086 and L11751. cf B_BR4.BZ-ID#.
4) **CM.CA-ID#:** These sequences are 3 of 17 sequences from a very diverse set of isolates from Yaounde and Douala, Cameroon. The sequences were derived from asymptomatic and symptomatic HIV infected individuals; specifically patients CA16 and CA20 were asymptomatic and patient CA4 was symptomatic. Virus was isolated by culture with donor PBMCs, and nested PCR amplified. A single clone was sequenced representing each HIV-1 isolate. The F subtype designation of these sequences is tentative. Although the F subtype sequences from Cameroon and Brazil consistently form a clade in phylogenetic analyses, the branch lengths between isolates from the two countries are typical of inter-subtype distances, and sequences from the two countries each form their own distinct clade within the F subtype (HIV database and Wouter Janssens, personal communication). Nkengasong JN, et al. *AIDS* **8**:1405–12 (1994). GenBank accession numbers for the entire set of 17 sequences studied in this publication: X80438–X80454.
5) **RO1.ID#:** These nine sequences are from isolates from Romanian children, in different clinical stages. All isolates showed cytopathic properties in peripheral blood mononuclear cells. They are also known as RM(A-J). Dumitrescu O, ML Kalish, SC Kliks, CI Bandea, JA Levy. *J Infect Dis* **169**:281–8 (1994). GenBank accession numbers L19571–L19579.
6) **RO2.RM-ID#:** These 24 sequences are from isolates from Romanian children. Unpublished: Holm-Hansen C, G Grothues, B Rosok, S Rustad, R Pascu and B Asjo, University of Bergen, Norway. GenBank accession numbers X77964–X77987.

G Subtype

At this time there are viral sequences from 11 HIV-1 infected individuals associated with HIV-1 subtype G. The G subtype consensus sequence (G_CONSENSUS_95) generated from these 8 sequences was based on the most common amino acid found in each position of an alignment. All 11 of these sequences have been published and/or have been made available for printing in the database by their authors.

1) **CF.4067:** This sequence was associated with the C subtype in first analysis of the C2V3 region (Murphy E, et al., *AIDS Res. Hum. Retroviruses* **9**:997–1006 (1993), but when a full gp120 sequence became avaliable from this isolate, and phylogenetic analysis was performed including some of the new subtype G sequences, it was more closely associated with G. The full length sequence was kindly provided prior to publication by Dr. MP Kieny of Transgene, Strasbourg Cedex, France. It is a part of a set of 14 HIV-1 gp120 isolates from Bangui, Central African Republic. The gp120 proteins have been expressed in a hybrid gp160 vaccinia virus expression system. Year of isolation and health status of individuals from which the virus was isolated were not provided. Viruses were isolated in the CAR by C Mathiot and B You (Pasteur Inst., Bangui), grown by F Barre-Sinoussi and A Deslandres (Pasteur Inst., Paris), and cloned and sequenced by D Schmitt and MP Kieny. GenBank accession numbers L11499 and L11500.
2) **GA.LBV21–7:** A sequence from Gabon from a set of HIV-1 viral isolates from Africa. This sequence was derived from a clone of PCR amplified DNA from cultured PBMCs. It represents a fragment of a full length env sequence. Janssens W, et al. *AIDS Res. Hum. Retroviruses* **10**:877–878 (1994). GenBank accession number U09664.
3) **GA.VI525:** A sequence from Gabon from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1994)] and Janssens W, et al. *AIDS Res. Hum. Retroviruses* **10**:877–878 (1994). GenBank accession number L22953.
4) **NG1.ID#:** These four sequences represent G subtype sequences from Nigeria. Abimiku AG, et al., *AIDS Res. Hum. Retroviruses* **10**:1581–1583 (1994). JP882 and JV832 were derived from AIDS patients, and G3 and G9 from healthy women. G9 was cultured on the T cell line CEM-SS, and the other three isolates were cocultured with uninfected donor PBMCs. DNA from viral cultures was PCR amplified, cloned and sequenced. GenBank accession numbers U13208–U13209, U13211 and U13213.
5) **NL.127C** This consensus sequence represents sequences generated from PCR amplified plasma RNA from one of three infants in a Dutch mother/infant study. A sample was collected from the infant at 1.5 months of age. Samples were also collected from the mother before birth, at birth and after birth. Mother sequences are not included in this consensus. [Mulder-Kampinga et al.(1993)]. [Mulder-Kampinga et al.(1995)]. Infant 127 sequences are from GenBank accession numbers Z47817–Z47832. Mother 127 sequences are from GenBank accession numbers Z47833–Z47880. Gag gene sequences from mother/child pairs are also available in Genbank accession numbers Z47903–Z47911; Z47912–Z47928; Z47929–Z47935; Z47936–Z47950. The second mother/child pair was also from the Netherlands, see B_NL.114C. The third mother/infant pair in this study was from Rwanda, see A_RW.564C.
6) **UG.K1:** This single sequence of G HIV1 is from an Ugandan patient. P. Kaleebu et al. (Unpublished) GenBank accession number U22010.
7) **UK1.22:** This sequence is a consensus of three clones from an infected infant in a mother-infant transmission study. The sequences were obtained via PCR from cell lysates, with sequencing of cloned PCR products. The infant was 3 months old at the time of blood drawing, and had pneumonia. [Arnold et al.(1995b)]. GenBank accession numbers U26304–U26306. Envelope sequences for the mother are found in genbank entries U26301–U26302, and gag sequences for mother and infant are in U26303 and U26307.
8) **??.ID#:** This sequence is from an unpublished GenBank entry by Lukashov,V.V. with accession number: L38413, RUS12A.

H Subtype

At this time there are viral sequences from 2 HIV-1 infected individuals associated with HIV-1 subtype H. The H subtype consensus sequence (H_CONSENSUS_95) generated from these 2 sequences was based on the most common amino acid found in each position of an alignment. Both of these sequences have been published and/or have been made available for printing in the database by their authors. Eight sequences that are too short for classification are closer to H than to other subtypes. The locus names (ID's) and sources of the sequences are:

1) **CM.CA13:** A sequence from Cameroon from a set of HIV-1 viral isolates from Africa used to define the prototype G and H env sequences. This sequence was derived from a clone of PCR amplified DNA from cultured PBMCs. It represents a fragment of a 900 base pair sequence. Janssens W, et al. *AIDS Res. Hum. Retroviruses* **10**:877–878 (1994) and Nkengasong JN, et al., *AIDS* **8**:1405–12 (1994). The H subtype association is not always clearly apparent using some sets of background sequences for comparison, and neighbor joining trees (HIV database, Wouter Janssens, personal communication), although parsimony trees confirmed the original association documented in Janssens W, et al. *AIDS Res. Hum. Retroviruses* **10**:877–878 (1994). GenBank accession number U09667.
2) **ZR.VI557:** A sequence from Zaire from a set of HIV-1 viral isolates from Africa used to define the prototype G and H env sequences. This sequence was derived from a clone of PCR amplified DNA from cultured PBMCs. It represents a fragment of a 900 base pair sequence. Janssens W, et al. *AIDS Res. Hum. Retroviruses* **10**:877–878 (1994). GenBank accession number U09666.

O Subtype

At this time there are viral sequences 12 HIV-1 infected individuals associated with HIV-1 subtype O that have been published and/or have been made available for printing in the database by their authors. The O subtype consensus sequence (O_CONSENSUS_95) generated from these 12 sequences was based on the most common amino acid found in each position of the alignment; when there was no consensus in a position an "X" was used. These sequences represent a set of sequences that are extremely divergent relative to other HIV-1's. The subtypes A-H have been grouped together under the heading "M" for main. "O" sequences are as different from one another as are sequences from different "M" subtypes. Only 3 of these sequences (ANT70, MVP5180 and VAU) were received in time to be included in the phenetic analysis.

1) **CM.CA9:** This sequence is from an individual living in Cameroon. W. Janssens et al. AIDS 8: 1012–1013 (1994). No GenBank entry is yet available. The pol gene from this isolate will be available in an entry with the GenBank accession number X78476.
2) **GA.VI686:** This sequence is from an individual living in Gabon. W. Janssens et al. AIDS 8: 1012–1013 (1994). No GenBank entry is yet available. The pol gene from this isolate will be available in an entry with the GenBank accession number X78477.
3) **CM.ANT70:** The complete viral genome has been sequenced from this viral isolate derived from a symptomatic Cameroonian, CDC stage III. R De Leys, et al., *J. Virol.* **64**:1207–1216 (1990) and V Haesevelde *J. Virol.*, **68**:1586–1596 (1994). GenBank accession number L20587.
4) **CM.MVP5180:** The complete viral genome has been sequenced from an isolate derived from a Cameroonian woman, sampled in 1991; the donor died of AIDS in 1992. The viral isolate MVP-5180 was grown in several human T-cell lines and the monocytic U937 line. Gurtler L, et al. *J. Virol.*, **68**:1581–5 (1994). GenBank accession number L20571.
5) **FR.VAU:** This sequence was derived from an isolate from a French woman who died of AIDS in 1992. DNA was extracted from VAU infected PBMCs, PCR amplified, cloned, and gp160 env was sequenced. The viral isolate was highly cytopathic. P Charneau et al., *Virology* 205: 247–253 (1994). GenBank accession number X80020.
6) **FR.CF#:** These seven consensus sequences are from Cameroonian patients living in France. I Loussert-Ajaka et al., J. Virol. 69: 5640–49 (1995). PBMC proviral DNA was PCR amplified and 3–6 clones from each patient were sequenced. The consensus of the 3–6 clones is presented. GenBank accession numbers U24562–U24568. Gag gene sequences for these patients are also available with GenBank accession numbers U24706–U24712.

Uncertain Classification

At this time there are viral sequences from 13 HIV-1 infected individuals that are not clearly associated with any of the HIV-1 genetic subtypes A through H. They either appeared distinct from the subtypes A-H in phylogenetic analysis, or else the subtype association was unclear, with different associations in different analyses. For some of the shorter gene fragments, subtype associations might have been established if more sequence information was available or if a different set of sequences was included in the background set used to define subtype associations. Some of these sequences may be representatives of subtypes as divergent as A-H, but only a single limited sample is yet available. Still others may represent recombinant genomes.

1) **BR3.RJI01** This sequence is B-F recombinant in the V3 region. DNA was amplified directly PBMCs of HIV infected individual, and the PCR product was directly sequenced. Sabino EC et al. *J. Virol.* **68**:6340–6346 (1994). GenBank accession number U00420.
2) **CF.ID#:** These four sequences are from a set of sequences obtained from 27 symptomatic patients from the Central African Republic, from whom blood was drawn in 1990–1991. Consensus, PCR-clones, cell culture, DNA. Murphy E, et al., *AIDS Res. Hum. Retroviruses* **9**:997–1006 (1993). GenBank accession numbers L11482–L11483, L11497, L11508–L11510 and L11514–L11515. Janssens et al. classified 4056 as an H subtype sequence (Janssens W, et al. *AIDS Res. Hum. Retroviruses* **10**:877–878 (1994)).
3) **KE.K124:** A Kenyan sequence from a set of HIV-1 viral isolates from Africa. Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. This isolate is not clearly associated with D subtype, however Louwagie and colleagues found that it associated with the D subtype in env, and the A subtype in gag. Using parsimony analysis, we found that it was difficult to determine a clear association, and this observation was confirmed by Wouter Janssens (personal communication). GenBank accession number L22942.
4) **KE.KEN976** This is a single unclassified sequence from a set of patients who were part of a 1990–1992 cohort study of maternal risk factors in mother to child transmission, including 22 pregnant women and an infant from Kenya. The C2V3 region was sequenced. W Janssens et al., in press *AIDS Res. Hum. Retroviruses* (1994). GenBank accession number U12992.
5) **NL.A11:** This sequence is from a Dutch study of presumed HIV-1 donor-recipient pairs. This sequence is from a recipient at the time of seroconversion; the donor was a Zairean woman living in the Netherlands. The sequences from both donor and recipient were extremely similar, so only the recipient is shown here. This sequence is a consensus sequences of multiple clones from PCR amplified serum RNA. Wolfs TFW, G Zwart, M Bakker, and J Goudmsit. *Virology* **189**:103–110 (1992). GenBank accession numbers M91849–M91856.
6) **ZM.ZAM184:** This Zambian sequence is an outlier, though in some phylogenetic analysis is appears most closely associated with the A subtype. In particular it is closely associated with A_CF.SAS (100/100 replicates in parsimony analysis of gp120). Health status of the individual from which the virus was cultured was unspecified, and the year of viral isolation was probably between 1989–1992. Viruses were cultured with donor PBMCs for 2 to 3 weeks. Full length env (gp160) was amplified, cloned and sequenced. [Louwagie et al.(1995)]. GenBank accession number L22955.
7) **ZR.Z3:** This sequence is from the 1983 Zairean isolate Z-3 (non-infectious, possibly due to frame-shift). Willey RW, et al., *Proc. Natl. Acad. Sci. USA* **83**:5038–5042 (1986). GenBank accession number K03347.

The three sequences which are not associated with any subtype and are not yet published, and for which the authors have not yet given permission for release, are from the following sets:

1) **ZR:** Two sequences from individuals from Zaire, provided by the LTCB, NCI, NIH, Bethesda, MD, USA (Dr. Marvin Reitz).

2) **GA:** A sequence provided by Dr. Wouter Janssens of the Prince Leopold Institute of Tropical Medicine, Antwerp, Belgium. It is a Gabonese isolate that is phylogenetically linked to F subtype sequences, but the relationship is distant.

References

[Ahmad et al.(1995)] N. Ahmad, B. M. Baroudy, R. C. Baker, & C. Chappey. Genetic analysis of human immunodeficiency virus type 1 envelope V3 region isolates from mothers and infants after perinatal transmission. *J Virol* **69**:1001–1012, 1995.

[Andeweg et al.(1992)] A. C. Andeweg, M. Groenink, P. Leeflang, R. E. de Goede, A. D. Osterhaus, M. Tersmette, & M. L. Bosch. Genetic and functional analysis of a set of HIV-1 envelope genes obtained from biological clones with varying syncytium-inducing capacities. *AIDS Res Hum Retroviruses* **8**:1803–13, 1992.

[Antonioli et al.(1995)] I. M. Antonioli, C. Baumberger, S. Yerly, & L. Perrin. V3 sequences in primary HIV-1 infection. *AIDS* **9**:11–17, 1995.

[Arnold et al.(1995a)] C. Arnold, P. Balfe, & J. P. Clewley. Sequence distances between env genes of HIV-1 from individuals infected from the same source: implications for the investigation of possible transmission events. *Virology* **211**:198–203, 1995a.

[Arnold et al.(1995b)] C. Arnold, K. L. Barlow, S. Kaye, P. Balfe, & J. P. Clewley. HIV type 1 sequence subtype G transmission from mother to infant: Failure of variant sequence species to amplify in the roche amplicor test. *AIDS Res Hum Retroviruses* **11**:999–1001, 1995b.

[Arnold et al.(1995c)] C. Arnold, K. L. Barlow, J. V. Parry, & J. P. Clewley. At least five HIV-1 sequence subtypes (A, B, C, D, A/E) occur in England. *AIDS Res Hum Retroviruses* **11**:427–429, 1995c.

[Artenstein et al.(1995)] A. W. Artenstein, T. C. VanCott, J. R. Mascola, J. K. Carr, P. A. Hegerich, J. Gaywee, E. Sanders-Buell, M. L. Robb, D. E. Dayhoff, S. Thitivichianlert, S. Nitayaphan, J. G. McNeil, D. L. Birx, R. A. Michael, D. S. Burke, & F. E. McCutchen. Dual infection with human immunodeficiency virus type 1 of distinct envelope subtypes in humans. *J Infect Dis* **171**:805–810, 1995.

[Bex et al.(1994)] F. Bex, P. Hermans, S. Sprecher, A. Achour, R. Badjou, C. Desgranges, J. Cogniaux, P. Franchioli, C. Vanhulle, & A. Lachgar. Syngeneic adoptive transfer of anti-human immunodeficiency virus (HIV-1)-primed lymphocytes from a vaccinated HIV-seronegative individual to his HIV-1-infected identical twin. *Blood* **8484**:3317–3326, 1994.

[Cabello et al.(1995)] A. Cabello, M. Cabral, M. Vera, R. Kiefer, R. Azorero, J. Eberle, L. Gurtler, & A. von Brunn. Analysis of the V3-loop sequences from 10 HIV-1 infected AIDS patients from Paraguay. *AIDS Res Human Retroviruses* **11**:1135–1137, 1995.

[Cheng-Mayer et al.(1991)] C. Cheng-Mayer, T. Shioda, & J. A. Levyt. Host range, replicative, and cytopathic properties of human immunodeficiency virus type 1 are determined by very few amino acid changes in tat and gp120. *J Virol* **65**:6931–6941, 1991.

[Cornelissen et al.(1995)] M. Cornelissen, G. Mulder-Campinga, J. Veenstra, F. Zorgdrager, C. Kuiken, S. Hartman, J. Dekker, L. V. D. Hoek, C. Sol, R. Coutinho, , & J. Goudsmit. Syncytium-inducing (SI) phenotype supression at seroconversion after intramuscular inoculation of a non-syncytium-inducing/SI phenotypically mixed human immunodeficiency virus population. *J Virol* **69**:1810–1818, 1995.

[Delwart et al.(1994)] E. L. Delwart, H. W. Sheppard, B. D. Walker, J. Goudsmit, & J. I. Mullins. Human immunodeficiency virus type 1 evolution in vivo tracked by DNA heteroduplex mobility assays. *J Virol* **68**:6672–6683, 1994.

[Diaz et al.(1995)] R. S. Diaz, E. C. Sabino, A. Mayer, J. W. Mosley, & M. P. Busch. Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. *J Virol* **69**:3273–3281, 1995.

[Fitzgibbon et al.(1993)] J. E. Fitzgibbon, S. Gaur, L. D. Frenkel, F. Laraque, B. R. Edlin, & D. T. Dubin. Transmission from one child to another of human immunodeficiency virus type 1 with a zidovudine-resistance mutation. *N Engl J Med* **329**:1835–41, 1993.

[Fouchier et al.(1992)] R. A. Fouchier, M. Groenink, N. A. Kootstra, M. Tersmette, H. G. Huisman, F. Miedema, & H. Schuitemaker. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. *J Virol* **66**:3183–7, 1992.

[Gao et al.(1994a)] F. Gao, L. Yue, S. Craig, C. L. Thornton, D. L. Robertson, F. E. McCutchan, J. A. Bradac, P. M. Sharp, B. H. Hahn, & the WHO Network for HIV Isolation and Characterization. Genetic variation of HIV type 1 in four World Health Organization-sponsored vaccine evaluation sites: generation of functional envelope (glycoprotein 160) clones representative of sequence subtypes A, B, C, and E. *AIDS Res Hum Retroviruses* **10**:1359–1368, 1994a.

[Gao et al.(1994b)] F. Gao, L. Yue, S. C. Hill, D. L. Robertson, A. H. Graves, M. S. Saag, G. M. Shaw, P. M. Sharp, & B. H. Hahn. HIV-1 sequence subtype D in the United States. *AIDS Res Hum Retroviruses* **10**:625–627, 1994b.

[Holmes et al.(1992)] E. C. Holmes, L. Q. Zhang, P. Simmonds, C. A. Ludlam, & A. J. Brown. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. *Proc Natl Acad Sci U S A* **89**:4835–9, 1992.

[Kuiken et al.(1995)] C. L. Kuiken, J. Goudsmit, G. F. Weiller, J. S. Armstrong, S. Hartman, P. Portegies, J. Dekker, & M. Cornelissen. Differences in human immunodeficiency virus type 1 V3 sequences from patients with and without AIDS dementia complex. *J Gen Virol* **76**:175–180, 1995.

[Louwagie et al.(1994)] J. Louwagie, E. L. Delwart, J. I. Mullins, F. E. McCutchan, G. Eddy, & D. S. Burke. Genetic analysis of HIV-1 isolates from Brazil reveals presence of two distinct genetic subtypes. *AIDS Res Hum Retroviruses* **10**:561–567, 1994.

[Louwagie et al.(1995)] J. Louwagie, W. Janssens, J. Mascola, L. Heyndrickx, P. Hegerich, G. van der Groen, F. E. McCutchan, & D. S. Burke. Genetic diversity of the envelope glycoprotein from human immunodeficiency virus type 1 isolates of African origin. *J Virol* **69**:263–271, 1995.

[Mascola et al.(1994)] J. R. Mascola, J. Louwagie, F. E. McCutchan, C. L. Fischer, P. A. Hegerich, K. F. Wagner, A. K. Fowler, J. G. McNeil, & D. S. Burke. Two antigenically distinct subtypes of human immunodeficiency virus type 1: viral genotype predicts neutralization serotype. *J Infect Dis* **169**:48–54, 1994.

[McNearney et al.(1993)] T. McNearney, Z. Hornickova, B. Kloster, A. Birdwell, G. A. Storch, M. A. S. H. Polmar, & L. Ratner. Evolution of sequence divergence among human immunodeficiency virus type 1 isolates derived from a blood donor and a recipient. *Pediatr Res* **33**:36–42, 1993.

[McNearney et al.(1992)] T. McNearney, Z. Hornickova, R. Markham, A. Birdwell, M. Arens, A. Saah, & L. Ratner. Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. *Proc Natl Acad Sci U S A* **89**:10247–10251, 1992.

[Monken et al.(1995)] C. E. Monken, B. Wu, & A. Srinivasan. High resolution analysis of HIV-1 quasispecies in the brain. *AIDS* **9**:345–349, 1995.

[Mulder-Kampinga et al.(1993)] G. A. Mulder-Kampinga, C. Kuiken, J. Dekker, H. J. Scherpbier, K. Boer, & J. Goudsmit. Genomic human immunodeficiency virus type 1 RNA variation in mother and child following intra-uterine virus transmission. *J Gen Virol* **74**:1747–56, 1993.

[Mulder-Kampinga et al.(1995)] G. A. Mulder-Kampinga, A. Simonon, C. L. Kuiken, J. Dekker, H. J. Scherpbier, P. van de Perre, K. Boer, & J. Goudsmit. Similarity in env and gag genes between genomic

RNAs of human immunodeficiency virus type 1 (HIV-1) from mother and infant is unrelated to time of HIV-1 RNA positivity in the child. *J. Virol.* **69**:2285–2296, 1995.

[Osmanov et al.(1994)] S. Osmanov, W. L. Heyward, & J. Esparza. The World Health Organization network for HIV isolation and characterization: summary of a pilot study. *AIDS Res Hum Retroviruses* **10**:1325–132, 1994.

[Ou et al.(1993)] C. Y. Ou, Y. Takebe, B. G. Weniger, C. C. Luo, M. L. Kalish, W. Auwanit, S. Yamazaki, N. L. Y. H. D. Gayle, & G. Schochetman. Independent introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand. *Lancet* **341**:1171–4, 1993.

[Pestano et al.(1995)] G. A. Pestano, K. Hosford, A. I. Spira, J. Riley, J.-M. Xie, N. Sewankambo, L. Brown, D. D. Ho, & W. M. O. Boto. Seroreactivity of analogous antigenic epitopes in glycoprotein 120 expressed in HIV-1 subtypes A, B, C, and D. *AIDS Res Hum Retroviruses* **11**:589–596, 1995.

[Reitz et al.(1994)] M. S. Reitz, L. Hall, M. Robert-Guroff, J. Lautenberger, B. M. Hahn, G. M. Shaw, L. I. Kong, S. H. Weiss, D. Waters, & R. C. Gallo. Viral variability and serum antibody response in a laboratory worker infected with HIV type 1 (HTLV type IIIB). *AIDS Res Hum Retroviruses* **10**:1143–1155, 1994.

[Sabino et al.(1994)] E. C. Sabino, E. Delwart, T. H. Lee, A. Mayer, J. I. Mullins, & M. P. Busch. Identification of low-level contamination of blood as basis for detection of human immunodeficiency virus (HIV) DNA in anti-HIV-negative specimens. *J Acquir Immune Defic Syndr* **7**:853–859, 1994.

[Sala et al.(1995)] M. Sala, E. Pelletier, & S. Wain-Hobson. HIV-1 gp120 sequences from a doubly infected drug user. *AIDS Res Hum Retroviruses* **11**:653–655, 1995.

[Sala et al.(1994)] M. Sala, G. Zambruno, J. P. Vartanian, A. Marconi, U. Bertazzoni, & S. Wain-Hobson. Spatial discontinuities in human immunodeficiency virus type 1 quasispecies derived from epidermal Langerhans cells of a patient with AIDS and evidence for double infection. *J Virol* **68**:5280–5283, 1994.

[Scarlatti et al.(1993)] G. Scarlatti, T. Leitner, E. Halapi, J. Wahlberg, P. Marchisio, M. A. Clerici-Schoeller, H. Wigzell, E. M. Fenyo, J.Albert, M. Uhlen, Rossi, & P. Comparison of variable region 3 sequences of human immunodeficiency virus type 1 from infected children with the rna and dna sequences of the virus populations of their mothers. *Proc Natl Acad Sci U S A* **90**:1721–5, 1993.

[Shapshak et al.(1995)] P. Shapshak, I. Nagano, K. Xin, W. Bradley, C. B. McCoy, N. C. Sun, R. V. Stewart, M. Yoshioka, C. Petito, K. Goodkin, & K. Okuda. HIV-1 heterogeneity and cytokines. neuropathogenesis. *Adv Exp Med Biol* **373**:225–238, 1995.

[Shpaer et al.(1994)] E. G. Shpaer, E. L. Delwart, C. L. Kuiken, J. Goudsmit, M. H. Bachmann, & J. I. Mullins. Conserved V3 loop sequences and transmission of human immunodeficiency virus type 1. *AIDS Res Hum Retroviruses* **10**:1679–1684, 1994.

[Siwka et al.(1994)] W. Siwka, A. Schwinn, K. Baczko, I. Pardowitz, F. Mhalu, J. Shao, A. Rethwilm, & V. ter Meulen. vpu and env sequence variability of HIV-1 isolates from Tanzania. *AIDS Res Hum Retroviruses* **10**:1753–1754, 1994.

[Slobod et al.(1994)] K. S. Slobod, S. D. Rencher, A. Farmer, F. S. Smith, & J. L. Hurwitz. HIV type 1 envelope sequence diversity in inner city community. *AIDS Res Hum Retroviruses* **10**:873–875, 1994.

[Srinivasan et al.(1987)] A. Srinivasan, R. Anand, D. York, P. Ranganathan, P. Feorino, G. Schochetman, J. Curran, V. S. Kalyanaraman, P. A. Luciw, & R. Sanchez-Pescador. Molecular characterization of human immunodeficiency virus from Zaire: nucleotide sequence analysis identifies conserved and variable domains in the envelope gene. *Gene* **52**:71–82, 1987.

[von Brunn et al.(1995)] A. von Brunn, B. von Brunn, J. Eberle, B. Biryahwaho, R. G. Downing, & L. Gurtler. Efficient antigen presentation to cytotoxic T lymphocytes by cells transduced with a retroviral vector expressing the HIV-1 Nef protein. *AIDS Res Hum Retroviruses* **11**:183–184, 1995.

[Wang et al.(1995)] W. K. Wang, M. Essex, & T. H. Leer. Uncommon gp120 cysteine residues found in primary HIV-1 isolates. *AIDS Res Hum Retroviruses* **11**:185–188, 1995.

[Wolf et al.(1994)] F. D. Wolf, E. Hogervorst, J. Goudsmit, E. M. Fenyo, H. Rubsamen-Waigmann, H. Holmes, B. Galvao-Castro, C. W. E. Karita, & S. D. Sempala. Syncytium-inducing and non-syncytium-inducing capacity of human immunodeficiency virus type 1 subtypes other than B: phenotypic and genotypic characteristics. *AIDS Res Hum Retroviruses* **10**:1387–1400, 1994.

[Wolfs et al.(1992)] T. F. Wolfs, G. Zwart, M. Bakker, & J. Goudsmit. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. *Virology* **189**:103–110, 1992.

[Wrin et al.(1989)] T. Wrin, T. P. Loh, J. C. Vennari, H. Schuitemaker, & J. H. Nunberg. Adaptation to persistent growth in the h9 cell line renders a primary isolate of human immunodeficiency virus type 1 sensitive to neutralization by vaccine sera. *J Virol* **69**:39–48, 1989.

[Xin et al.(1995)] K. Q. Xin, P. Shapshak, S. Kawamoto, I. Nagano, C. B. McCoy, & K. Okuda. Highly divergent env sequences of HIV-1 B subtype with two novel V3 loop motifs detected in an AIDS patient in Miami, Florida. *AIDS Res Hum Retroviruses* **11**:1139–1141, 1995.

[Yu et al.(1995)] X. F. Yu, Z. Wang, C. Beyrer, D. D. Celentano, C. Khamboonruang, E. Allen, & K. Nelson. Phenotypic and genotypic characteristics of human immunodeficiency virus type 1 from patients with AIDS in northern Thailand. *J Virol* **69**:4649–4655, 1995.

[Zhang et al.(1993)] L. Q. Zhang, P. MacKenzie, A. Cleland, E. C. Holmes, A. J. Brown, & P. Simmonds. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. *J Virol* **67**:3345–3356, 1993.

[Zhu et al.(1995)] T. Zhu, N. Wang, A. Carr, S. Wolinsky, & D. D. Ho. Evidence for coinfection by multiple strains of human immunodeficiency virus type 1 subtype B in an acute seroconvertor. *J Virol* **69**:1324–1327, 1995.